ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

ITERATION OF ANALYTIC FUNCTIONS

BY CARL LUDWIG SIEGEL

(Received April 23, 1942)

Let

$$f(z) = \sum_{k=1}^{\infty} a_k z^k \tag{1}$$

be a power series without constant term and denote by $R > 0$ its radius of convergence. The fixed point $z = 0$ of the mapping $z \to f(z)$ is called stable, if there exist two positive finite numbers $r_0 \leqq R$ and $r \leqq R$, such that for all points z of the circle $|z| < r_0$ the set of image points $z_1 = f(z)$, $z_{n+1} = f(z_n)$ $(n = 1, 2, \cdots)$ lies in the circle $|z| < r$.

It is easy to prove the stability in the case $|a_1| < 1$, for then a positive number $r_0 < R$ exists, such that the inequality $|f(z)| \leqq |z|$ holds for $|z| < r_0$, and $r = r_0$ has the required property. Henceforth, the inequality $|a_1| \geqq 1$ is assumed.

If $z = 0$ is stable, then the images z_n $(n = 1, 2, \cdots)$ of the points z of the circle $|z| < r_0$ under the mapping $z \to f(z)$ and its iterations cover a domain D which is connected and contains the point $z = 0$. For all z in D, the image point $f(z)$ again lies in D. Let

$$z = \varphi(\zeta) = \zeta + \sum_{k=2}^{\infty} c_k \zeta^k \tag{2}$$

be the power series mapping a certain circle $|\zeta| < \rho$ of the ζ plane conformally onto the universal covering surface of D. Then the formula

$$\varphi(\zeta) = z \to f(z) = z_1 = \varphi(\zeta_1)$$

defines a function $\zeta_1 = g(\zeta)$ which is regular in the circle $|\zeta| < \rho$ and satisfies there the inequality $|g(\zeta)| < \rho$; moreover $g(0) = 0$ and $g'(0) = 1$. It follows from Schwarz's lemma that $|a_1| = 1$ and $\zeta_1 = a_1\zeta$. Consequently, the functional equation of Schröder

$$\varphi(a_1\zeta) = f(\varphi(\zeta)) \tag{3}$$

has a convergent solution $\varphi(\zeta) = \zeta + \cdots$.

On the other hand, it is obvious that $z = 0$ is stable, if $|a_1| = 1$ and the functional equation (3) has a convergent solution.

If a_1 is an n^{th} root of unity, then $z = 0$ is stable, if and only if the $(n-1)^{\text{th}}$ iteration of the mapping $z \to f(z)$ is the identity. This is also easily proved by direct calculation. We assume now that $|a_1| = 1$ and $a_1^n \neq 1$ for $n = 1, 2, \cdots$.

By (1), (2) and (3),

$$\sum_{k=2}^{\infty} c_k(a_1^k - a_1)\zeta^k = \sum_{l=2}^{\infty} a_l \left(\zeta + \sum_{r=2}^{\infty} c_r \zeta^r\right)^l; \tag{4}$$

hence c_k $(k = 2, 3, \cdots)$ is a polynomial in $c_2, \cdots, c_{k-1}$ whose coefficients depend upon $a_1, \cdots, a_k$, and there exists exactly one formal (convergent or divergent) solution $\varphi(\zeta) = \zeta + \cdots$ of (3). The first example of a convergent series $f(z) = a_1 z + \cdots$ with *divergent* Schröder series $\varphi(\zeta)$ has been given by Pfeiffer.[1] Later Cremer[2] has constructed such examples for arbitrary a_1 satisfying the condition

$$\liminf_{n \to \infty} |a_1^n - 1|^{1/n} = 0.$$

These a_1 are very closely approximated by certain roots of unity, and their linear Lebesgue measure on the unit circle $|a_1| = 1$ is 0.

Until now, however, it was not known if there exists a number a_1 of absolute value 1, such that every convergent power series $f(z) = a_1 z + \cdots$ has a *convergent* Schröder series. Julia[3] tried to prove the erroneous hypothesis that the Schröder series is always divergent, if $f(z) - a_1 z$ is a rational function and not identically 0. We shall demonstrate the following

THEOREM: *Let*

$$\log |a_1^n - 1| = O(\log n) \qquad (n \to \infty); \tag{5}$$

then the Schröder series is convergent.

Write $a_1 = e^{2\pi i \omega}$; then the condition (5) may be expressed in the form

$$\left| \omega - \frac{m}{n} \right| > \lambda n^{-\mu},$$

for arbitrary integers m and n, $n \geqq 1$, where λ and μ denote positive numbers depending only upon ω. It is easily seen that (5) holds for all points of the unit circle $|a_1| = 1$ with the exception of a set of measure 0.

LEMMA 1: *Let x_p $(p = 1, \cdots, r)$ and y_q $(q = 1, \cdots, s)$ be positive integers, $r \geqq 0$, $s \geqq 2$, $r + s = t$,*

$$\sum_{p=1}^{r} x_p + \sum_{q=1}^{s} y_q = k, \qquad \sum_{q=1}^{s} y_q > \frac{k}{2}, \qquad y_q \leqq \frac{k}{2} \ (q = 1, \cdots, s);$$

then

$$\prod_{p=1}^{r} x_p \prod_{q=1}^{s} y_q^2 \geqq k^3 8^{1-t}. \tag{6}$$

PROOF: Denote the left-hand side of (6) by L and consider first the case $k < 2t - 2$. Then

$$k^{-3} L \geqq k^{-3} > (2t - 2)^{-3}. \tag{7}$$

---

[1] G. A. Pfeiffer, *On the conformal mapping of curvilinear angles. The functional equation* $\varphi[f(x)] = a_1 \varphi(x)$, Trans. Amer. Math. Soc. 18, pp. 185–198 (1917).

[2] H. Cremer, *Über die Häufigkeit der Nichtzentren*, Math. Ann. 115, pp. 573–580 (1938).

[3] G. Julia, *Sur quelques problèmes relatifs à l'itération des fractions rationnelles*, C. R. Acad. Sci. Paris 168, pp. 147–149 (1919).

Assume now $k \geqq 2t - 2$ and let

$$\left[\frac{k}{2}\right] = g, \qquad r + \sum_{q=1}^{s} y_q = \eta.$$

Then

$$t \leqq g + 1 \leqq g + 1 + r \leqq \eta \leqq k, \qquad \sum_{p=1}^{r} x_p = k - \eta + r,$$

whence

$$\prod_{p=1}^{r} x_p \geqq k - \eta + 1, \quad \prod_{q=1}^{s} y_q \geqq \begin{cases} \eta - t + 1, & \text{if } \eta \leqq g - 1 + t \\ (\eta - g - t + 2)g, & \text{if } \eta \geqq g - 1 + t. \end{cases}$$

In the interval $g + 1 \leqq \eta \leqq g - 1 + t$,

$$(k - \eta + 1)(\eta - t + 1)^2 \geqq \min \{(k - g)(g - t + 2)^2, (k - g - t + 2)g^2\};$$

in the interval $g - 1 + t \leqq \eta \leqq k$,

$$(k - \eta + 1)(\eta - g - t + 2)^2 g^2 \geqq (k - g - t + 2)g^2;$$

in the interval $0 \leqq \xi \leqq g$,

$$(k - g)(g - \xi)^2 - (k - g - \xi)g^2 = \{(k - g)\xi - (2k - 3g)g\}\xi \leqq g(2g - k)\xi \leqq 0;$$

consequently

$$L \geqq (k - g)(g - t + 2)^2$$

$$\text{(8)} \qquad k^{-3} L \geqq \frac{k - g}{k}\left(\frac{g - t + 2}{k}\right)^2 \geqq \tfrac{1}{2}(2t - 2)^{-2} \geqq (2t - 2)^{-3}.$$

Now

$$t - 1 \leqq 2^{t-2} \qquad (t = 2, 3, \cdots),$$

and the lemma follows from (7) and (8).

We use the abbreviation

$$\epsilon_n = |\, a_1^n - 1\,|^{-1} \qquad (n = 1, 2, \cdots).$$

On account of (5), the inequalities

$$\epsilon_n < (2n)^{\nu} \qquad (n = 1, 2, \cdots)$$

are fulfilled for a certain constant positive value ν. We define

$$N_1 = 2^{2\nu+1}, \qquad N_2 = 8^{\nu} N_1 = 2^{5\nu+1}.$$

LEMMA 2: *Let m_l $(l = 0, \cdots, r)$ be integral, $r \geqq 0$ and $m_0 > m_1 > \cdots > m_r > 0$; then*

$$\text{(9)} \qquad \prod_{l=0}^{r} \epsilon_{m_l} < N_1^{r+1}\left\{m_0 \prod_{l=1}^{r} (m_{l-1} - m_l)\right\}^{\nu}.$$

Proof: The assertion is true in the case $r = 0$; assume $r > 0$ and apply induction.

We have the identity

$$a_1^q(a_1^{p-q} - 1) = (a_1^p - 1) - (a_1^q - 1) \qquad (0 < q < p),$$

whence

$$\epsilon_{p-q}^{-1} \leqq \epsilon_p^{-1} + \epsilon_q^{-1}$$

$$\min(\epsilon_p, \epsilon_q) \leqq 2\epsilon_{p-q} < 2^{\nu+1}(p-q)^\nu.$$

This simple remark is the main argument of the whole proof.

Let ϵ_{m_l} $(l = 0, \cdots, r)$ have its minimum value for $l = h$. Then

$$\epsilon_{m_h} < 2^{\nu+1}\min\{(m_{h-1} - m_h)^\nu, (m_h - m_{h+1})^\nu\}, \tag{10}$$

if we define moreover $m_{-1} = \infty$ and $m_{r+1} = -\infty$. On the other hand, the lemma being true for $r - 1$ instead of r, we have

$$\epsilon_{m_h}^{-1}\prod_{l=0}^{r}\epsilon_{m_l} < N_1^r\left\{\frac{m_0(m_{h-1} - m_{h+1})}{(m_{h-1} - m_h)(m_h - m_{h+1})}\prod_{l=1}^{r}(m_{l-1} - m_l)\right\}^\nu. \tag{11}$$

Since

$$\frac{m_{h-1} - m_{h+1}}{(m_{h-1} - m_h)(m_h - m_{h+1})} = \frac{1}{m_{h-1} - m_h} + \frac{1}{m_h - m_{h+1}} \leqq \frac{2}{\min(m_{h-1} - m_h, m_h - m_{h+1})},$$

the inequality (9) follows from (10) and (11).

Consider now the sequence of positive numbers $\delta_1 = 1, \delta_2, \delta_3, \cdots$ recurrently defined in the following way: For every $k > 1$, let μ_k denote the maximum of all products $\delta_{l_1}\delta_{l_2}\cdots\delta_{l_r}$ with $l_1 + l_2 + \cdots + l_r = k > l_1 \geqq l_2 \geqq \cdots \geqq l_r \geqq 1$, $2 \leqq r \leqq k$, and put

$$\delta_k = \epsilon_{k-1}\mu_k. \tag{12}$$

Lemma 3:

$$\delta_k \leqq k^{-2\nu}N_2^{k-1} \qquad (k = 1, 2, \cdots). \tag{13}$$

Proof: The assertion is true in the case $k = 1$; assume $k > 1$ and apply induction.

The numbers $\alpha_k = k^{-2\nu}N_2^{k-1}$ satisfy the inequalities

$$\frac{\alpha_k\alpha_l}{\alpha_{k+l}} = (k^{-1} + l^{-1})^{2\nu}N_2^{-1} \leqq 2^{2\nu}N_2^{-1} < 1 \qquad (k \geqq 1, l \geqq 1),$$

and consequently

$$\delta_{j_1}\delta_{j_2}\cdots\delta_{j_f} \leqq j^{-2\nu}N_2^{j-1} \qquad (1 \leqq j_1 + \cdots + j_f = j < k; f \geqq 1). \tag{14}$$

By (12), there exists a decomposition

$$\delta_k = \epsilon_{k-1}\delta_{g_1}\delta_{g_2}\cdots\delta_{g_\alpha} \qquad (g_1 + \cdots + g_\alpha = k > g_1 \geqq \cdots \geqq g_\alpha \geqq 1).$$

In the case $g_1 > k/2$, we use this formula with g_1 instead of k and find a decomposition

$$\delta_{g_1} = \epsilon_{g_1-1}\delta_{h_1}\delta_{h_2} \cdots \delta_{h_\beta} \qquad (h_1 + \cdots + h_\beta = g_1 > h_1 \geqq \cdots \geqq h_\beta \geqq 1);$$

if also $h_1 > k/2$, we decompose again

$$\delta_{h_1} = \epsilon_{h_1-1}\delta_{i_1}\delta_{i_2} \cdots \delta_{i_\gamma} \qquad (i_1 + \cdots + i_\gamma = h_1 > i_1 \geqq \cdots \geqq i_\gamma \geqq 1),$$

and so on. Writing $k_0 = k$, $k_1 = g_1$, $k_2 = h_1$, $\cdots$, we obtain in this manner the formula

$$\delta_k = \prod_{p=0}^{r} (\epsilon_{k_p-1}\Delta_p)$$

with $k = k_0 > k_1 > \cdots > k_r > k/2$, where Δ_p denotes for $p = 0, \cdots, r$ a certain product $\delta_{j_1} \cdots \delta_{j_f}$ and

$$j_1 + \cdots + j_f = \begin{cases} k_p - k_{p+1} & (p = 0, \cdots, r-1) \\ k_r & (p = r), \end{cases}$$

all subscripts $j_1, \cdots, j_f$ being $\leqq k/2$. The number f depends upon p; let $f = s$ for $p = r$.

Using (13) for the s single factors of Δ_r and applying (14) for the estimation of Δ_p ($p = 0, \cdots, r-1$), we find the inequality

$$\prod_{p=0}^{r} \Delta_p \leqq N_2^{k-r-s} \left\{ \prod_{q=1}^{s} j_q \prod_{p=1}^{r} (k_{p-1} - k_p) \right\}^{-2\nu},$$

where $1 \leqq j_q \leqq k/2$ ($q = 1, \cdots, s$) and $j_1 + \cdots + j_s = k_r$. By Lemma 2,

$$\prod_{p=0}^{r} \epsilon_{k_p-1} < N_1^{r+1} \left\{ k \prod_{p=1}^{r} (k_{p-1} - k_p) \right\}^{\nu},$$

and consequently

$$\delta_k < N_1^{r-1} N_2^{k-t} \left(k^{-1} \prod_{p=1}^{r} x_p \prod_{q=1}^{s} y_q^2 \right)^{-\nu}$$

with $t = r + s$, $x_p = k_{p-1} - k_p$, $y_q = j_q$. By Lemma 1,

$$N_2^{1-k} k^{2\nu} \delta_k < N_1^{r+1} N_2^{1-t} 8^{\nu(t-1)} \leqq \left(\frac{8^\nu N_1}{N_2} \right)^{t-1} = 1,$$

and (13) is proved.

Proof of the Theorem: Since the power series (1) has a positive radius of convergence, there exists a positive number a, such that $|a_n| \leqq a^{n-1}$ ($n = 2, 3, \cdots$). The functional equation (3) remains true under the transformation $f(z) \to af(z/a)$, $\varphi(\zeta) \to a\varphi(\zeta/a)$; hence we may assume $|a_n| \leqq 1$ ($n = 2, 3, \cdots$).

Instead of (4), we consider the functional equation

$$\sum_{k=2}^{\infty} \eta_k \gamma_k \zeta^k = \sum_{l=2}^{\infty} \left(\zeta + \sum_{r=2}^{\infty} \gamma_r \zeta^r \right)^l, \tag{15}$$

where $\eta_2, \eta_3, \cdots$ are positive parameters. Then the coefficients $\gamma_1 = 1, \gamma_2, \gamma_3, \cdots$ are uniquely determined by the formula

$$\gamma_k = \eta_k^{-1} \sum \gamma_{l_1}\gamma_{l_2} \cdots \gamma_{l_r} \qquad (k = 2, 3, \cdots), \tag{16}$$

where $l_1, \cdots, l_r$ run over all positive integral solutions of $l_1 + \cdots + l_r = k$ $(r = 2, \cdots, k)$. Write $\gamma_k = \sigma_k$ in the case $\eta_k = \epsilon_{k-1}^{-1}$ $(k = 2, 3, \cdots)$, and $\gamma_k = \tau_k$ in the case $\eta_k = 1$.

The inequality

$$\sigma_k \leqq \delta_k \tau_k \tag{17}$$

is true for $k = 1$. Applying induction, we infer from (12) and (16) that

$$\sigma_k \leqq \epsilon_{k-1}\mu_k \sum \tau_{l_1}\tau_{l_2} \cdots \tau_{l_r} = \delta_k \tau_k ;$$

hence (17) holds for all values of k.

On the other hand, the power series

$$\psi = \sum_{k=1}^{\infty} \tau_k \zeta^k$$

satisfies the equation

$$\psi - \zeta = (1 - \psi)^{-1}\psi^2,$$

whence

$$4\psi = 1 + \zeta - (1 - 6\zeta + \zeta^2)^{\frac{1}{2}};$$

consequently ψ converges in the circle $|\zeta| < 3 - 2\sqrt{2}$.

By (4), (15) and (17),

$$|c_k| \leqq \delta_k \tau_k \qquad (k = 2, 3, \cdots).$$

It follows now from Lemma 3, that the Schröder series $\varphi(\zeta)$ converges in the circle $|\zeta| < (3 - 2\sqrt{2})2^{-5\nu-1}$.

INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

NOTE ON AUTOMORPHIC FUNCTIONS OF SEVERAL VARIABLES

BY CARL LUDWIG SIEGEL

(Received April 28, 1942)

1

Some years ago I found a method[1] of estimating the number of linearly independent modular forms of degree n and of weight g, which has been useful for the demonstration[2] of certain identities in the analytical theory of quadratic forms. The object of this note is to prove an analogous estimate concerning automorphic functions.

Let $\mathfrak{Z} = (z_{kl})$ be a complex symmetric matrix with n rows, and consider the space E defined by the condition $\mathfrak{E} - \mathfrak{Z}\bar{\mathfrak{Z}} > 0$, with the line element

$$ds = \sigma^{\frac{1}{2}}\{d\mathfrak{Z}(\mathfrak{E} - \bar{\mathfrak{Z}}\mathfrak{Z})^{-1}d\bar{\mathfrak{Z}}(\mathfrak{E} - \mathfrak{Z}\bar{\mathfrak{Z}})^{-1}\},$$

the symbol σ denoting the trace. If $\mathfrak{A}$ and $\mathfrak{B}$ are n-rowed complex square matrices satisfying $\mathfrak{A}\mathfrak{B}' = \mathfrak{B}\mathfrak{A}'$ and $\mathfrak{A}\bar{\mathfrak{A}}' - \mathfrak{B}\bar{\mathfrak{B}}' = \mathfrak{E}$, then the linear transformation

$$\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\bar{\mathfrak{B}}\mathfrak{Z} + \bar{\mathfrak{A}})^{-1} \tag{1}$$

defines an isometric mapping of E onto itself. Those transformations constitute a group Ω.

Denoting by $\rho(\mathfrak{Z}_1, \mathfrak{Z}_0)$ the distance of two arbitrary points $\mathfrak{Z}_1$ and $\mathfrak{Z}_0$ of E, we have[3]

$$\rho(\mathfrak{Z}_1, 0) = \left(\sum_{k=1}^{n} u_k^2\right)^{\frac{1}{2}},$$

where

$$u_k = \log \frac{1 + \lambda_k^{\frac{1}{2}}}{1 - \lambda_k^{\frac{1}{2}}} \qquad (k = 1, \cdots, n)$$

and $\lambda_1, \cdots, \lambda_n$ are the characteristic roots of the hermitian matrix $\mathfrak{Z}_1\bar{\mathfrak{Z}}_1$. Since

$$\frac{4\lambda_k}{1 - \lambda_k} = e^{u_k} + e^{-u_k} - 2 = 2\sum_{l=1}^{\infty} \frac{u_k^{2l}}{(2l)!}$$

---

[1] C. L. Siegel, *Einführung in die Theorie der Modulfunktionen n-ten Grades*, Math. Ann. 116, pp. 617–657 (1939).

[2] H. Maass, *Zur Theorie der automorphen Funktionen von n Veränderlichen*, Math. Ann. 117, pp. 538–578 (1940).

E. Witt, *Eine Identität zwischen Modulformen zweiten Grades*, Abh. Math. Sem. Hansischen Univ. 14, pp. 323–337 (1941).

H. Maass, *Modulformen und quadratische Formen über dem quadratischen Zahlkörper $R(\sqrt{5})$*, Math. Ann. 118, pp. 65–84 (1942).

[3] C. L. Siegel, *Symplectic geometry*, submitted for publication in the Amer. J. Math.

and

$$\sum_{k=1}^{n} u_k^{2l} \leqq \left(\sum_{k=1}^{n} u_k^2\right)^l = \rho^{2l}(\mathfrak{Z}_1, 0),$$

we obtain the inequality

$$\sum_{k=1}^{n} \frac{\lambda_k}{1-\lambda_k} \leqq \sinh^2 \tfrac{1}{2}\rho, \qquad \rho = \rho(\mathfrak{Z}_1, 0). \tag{2}$$

Let Δ be a subgroup of Ω, discontinuous in E, and assume that all frontier points of a fundamental domain F of Δ belong to E; i.e. E is compact relative to Δ. The least upper bound of the distance $\rho(\mathfrak{Z}_1, \mathfrak{Z}_0)$ for two variable points $\mathfrak{Z}_1$ and $\mathfrak{Z}_0$ of F is a finite positive number δ, the diameter of F. We use the abbreviations

$$\nu = \frac{n(n+1)}{2}, \qquad b = \sinh^2 \tfrac{1}{2}\delta, \qquad c = (\nu+1)b^\nu. \tag{3}$$

2

An analytic function $f(\mathfrak{Z})$ of the ν independent variables z_{kl} $(1 \leqq k \leqq l \leqq n)$ is called an automorphic form with the group Δ, if it is regular in E and satisfies there the equations

$$f(\mathfrak{Z}^*) = v(\mathfrak{A}, \mathfrak{B})\, |\, \mathfrak{B}\mathfrak{Z} + \bar{\mathfrak{A}}\, |^{-g} f(\mathfrak{Z}) \tag{4}$$

for all transformations (1) in the group Δ, where g is a constant and the numbers $v = v(\mathfrak{A}, \mathfrak{B})$ depend only upon $\mathfrak{A}$ and $\mathfrak{B}$. Let $L = L(\Delta, g, v)$ denote the set of all such functions $f(\mathfrak{Z})$, the weight g and the multiplier system v being given. If f_1 and f_2 belong to this set, then so does $\lambda f_1 + \mu f_2$, for arbitrary complex constants λ and μ; hence L is a vector space with a certain (finite or infinite) dimension d.

For automorphic forms of a single variable, i.e. in the case $n = 1$, the number d is given by the generalized Riemann-Roch theorem.[4] It is not known in which way this theorem might be extended to automorphic forms of several variables. We now assume that the weight g is real and that all multipliers $v(\mathfrak{A}, \mathfrak{B})$ have absolute value 1. We shall derive a finite upper bound of d depending only upon n, g and δ.

Consider first the case $g = 0$. Then, by (4) the absolute value abs $f(\mathfrak{Z})$ is invariant under Δ; consequently it attains in E a maximum at an inner point. This proves $f(\mathfrak{Z})$ is a constant, whence $d = 1$, if $v(\mathfrak{A}, \mathfrak{B}) = 1$, and $d = 0$ otherwise. In the remainder of the paper, we suppose $g \neq 0$.

LEMMA: *Let $f(\mathfrak{Z})$ be a function of the set $L(\Delta, g, v)$, not identically* 0. *If all its partial derivatives of the orders* $0, 1, \cdots, h-1$ $(h \geqq 0)$ *vanish at a point $\mathfrak{Z}_0$ of E, then $h \leqq bg$.*

---

[4] E. Ritter, *Die multiplicativen Formen auf algebraischem Gebilde beliebigen Geschlechtes mit Anwendung auf die Theorie der automorphen Formen*, Math. Ann. 44, pp. 261–374 (1894).
H. Petersson, *Zur analytischen Theorie der Grenzkreisgruppen, Teil II*, Math. Ann. 115, pp. 175–204 (1938).

PROOF: The continuous function

$$\varphi(\mathfrak{Z}) = |\,\mathfrak{E} - \mathfrak{Z}\bar{\mathfrak{Z}}\,|^{\frac{1}{2}g} \operatorname{abs} f(\mathfrak{Z})$$

is invariant under Δ; consequently it has in E a maximum $\mu > 0$, which is attained at a point $\mathfrak{Z}_1$ of F. On account of (4), we may assume that $\mathfrak{Z}_0$ also lies in F. In case $h > 0$, the function $f(\mathfrak{Z})$ vanishes at $\mathfrak{Z} = \mathfrak{Z}_0$, whence $\mathfrak{Z}_1 \neq \mathfrak{Z}_0$. In case $h = 0$, the assumption of the lemma holds for every point $\mathfrak{Z}_0$ of E, and we may suppose $\mathfrak{Z}_1 \neq \mathfrak{Z}_0$.

If the transformation (1) is any given element M of the group Ω, then the function $|\,\bar{\mathfrak{B}}\mathfrak{Z} + \bar{\mathfrak{A}}\,|^{-g} f(\mathfrak{Z}^*)$ belongs to $L(M^{-1}\Delta M, g, v)$. Since Ω is transitive in E and the diameter δ is invariant under Ω, we may assume for the proof of the lemma that $\mathfrak{Z}_0 = 0$ and $\rho(\mathfrak{Z}_1, 0) \leqq \delta$. Let $\lambda_1, \cdots, \lambda_n$ be the characteristic roots of $\mathfrak{Z}_1\bar{\mathfrak{Z}}_1$, $0 \leqq \lambda_1 \leqq \cdots \leqq \lambda_n$; then $0 < \lambda_n < 1$ and, by (2) and (3),

$$0 < \sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_k} \leqq b. \tag{5}$$

We introduce a single complex variable z and choose in particular $\mathfrak{Z} = z\mathfrak{Z}_1$. For all points z of the circle $z\bar{z} < \lambda_n^{-1}$, the matrix $\mathfrak{Z}$ lies in E; hence there $f(\mathfrak{Z})$ is a regular analytic function $\psi(z)$ which vanishes at the point $z = 0$ at least of the order h and satisfies the relationship

$$\operatorname{abs} \psi(z) = |\,\mathfrak{E} - z\bar{z}\mathfrak{Z}_1\bar{\mathfrak{Z}}_1\,|^{-\frac{1}{2}g}\varphi(z\mathfrak{Z}_1) \leqq |\,\mathfrak{E} - z\bar{z}\mathfrak{Z}_1\bar{\mathfrak{Z}}_1\,|^{-\frac{1}{2}g}\mu,$$

where the equality holds for $z = 1$.

Let $1 < t < \lambda_n^{-1}$. On the circle $z\bar{z} \leqq t$, the analytic function $z^{-h}\psi(z)$ attains the maximum of its absolute value at a point of the boundary, whence

$$\operatorname{abs} \psi(1) \leqq t^{-\frac{1}{2}h} \max_{z\bar{z}=t} \operatorname{abs} \psi(z)$$

$$|\,\mathfrak{E} - \mathfrak{Z}_1\bar{\mathfrak{Z}}_1\,|^{-\frac{1}{2}g}\mu \leqq t^{-\frac{1}{2}h}\,|\,\mathfrak{E} - t\mathfrak{Z}_1\bar{\mathfrak{Z}}_1\,|^{-\frac{1}{2}g}\mu.$$

But $|\,\mathfrak{E} - t\mathfrak{Z}_1\bar{\mathfrak{Z}}_1\,| = \prod_{k=1}^{n}(1 - t\lambda_k)$ and therefore

$$h \leqq g \log \prod_{k=1}^{n} \frac{1 - \lambda_k}{1 - t\lambda_k} \Big/ \log t \qquad (1 < t < \lambda_n^{-1}).$$

Performing the passage to the limit $t \rightarrow 1$, we obtain the inequality

$$h \leqq g \sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_k}. \tag{6}$$

The assertion of the lemma follows from (5) and (6).

THEOREM: *The dimension d of $L(\Delta, g, v)$ is 0 for $g < b^{-1}$ and $\leqq cg^n$ for $g > 0$.*

PROOF: Assume $d > 0$ and choose in $L(\Delta, g, v)$ a function $f(\mathfrak{Z})$, which does not vanish identically. Applying the lemma with $h = 0$, we infer $0 \leqq bg$. This proves the theorem in the case $g < 0$.

Now consider the case $g > 0$. If $f(\mathfrak{Z}) \neq 0$ everywhere in E, then $f^{-1}(\mathfrak{Z})$ is a non-vanishing function of the set $L(\Delta, -g, v^{-1})$ and $-g < 0$, which is impossible. Consequently we may apply the lemma with $h = 1$ and obtain the

inequality $1 \leqq bg$, whence $1 < (\nu + 1)(bg)^\nu = cg^\nu$. This proves the theorem in the case $g > 0$ and $d = 0$ or 1.

In the remaining case $g > 0, d \geqq 2$, let $f_1, \cdots, f_m$ be a finite number of linearly independent functions in $L(\Delta, g, v)$ and $m \geqq 2$. We determine the positive integer h by the condition

$$\binom{\nu + h - 1}{\nu} < m \leqq \binom{\nu + h}{\nu} \tag{7}$$

and choose m constants $a_1, \cdots, a_m$, not all 0, such that all partial derivatives of the orders $0, 1, \cdots, h - 1$ vanish for the function

$$f(\mathfrak{Z}) = a_1 f_1 + \cdots + a_m f_m$$

at the point $\mathfrak{Z} = 0$; this is possible, by (7), since we have to satisfy $\binom{\nu + h - 1}{\nu}$ homogeneous linear equations with the m unknown quantities $a_1, \cdots, a_m$. By (7) and the lemma,

$$m \leqq \binom{\nu + h}{\nu} \leqq (\nu + 1)h^\nu \leqq (\nu + 1)(bg)^\nu = cg^\nu.$$

This proves the remaining part of the theorem.

3

A function $\chi(\mathfrak{Z})$ is called an automorphic function with the group Δ, if $\chi(\mathfrak{Z}) = f_1/f_0$, f_0 not identically 0, where f_1 and f_0 are automorphic forms in the same set $L(\Delta, g, v)$. For a sufficiently large value $G > 0$, certain functions in the set $L(\Delta, G, 1)$ can be expressed as Poincaré series,[5] and it may be proved by known methods that there exist $\nu + 1$ of those functions, say $F_0, \cdots, F_\nu$, which are algebraically independent. Then the ν quotients $\chi_k = F_k/F_0$ $(k = 1, \cdots, \nu)$ are algebraically independent automorphic functions with the group Δ.

Define $q = [c\nu! G^\nu]$ and choose a positive integer Q satisfying the condition $q + 1 > c\nu!(G + gqQ^{-1})^\nu$. The number of power products

$$P = \chi^r \prod_{k=1}^{\nu} \chi_k^{s_k}$$

with $0 \leqq r \leqq q$, $0 \leqq s_k$ $(k = 1, \cdots, \nu)$, $s_1 + \cdots + s_\nu \leqq Q$ is

$$A = (q + 1)\binom{Q + \nu}{\nu} > \frac{q + 1}{\nu!} Q^\nu > c(gq + GQ)^\nu; \tag{8}$$

we denote them by $P_1, \cdots, P_A$. Then the A functions $f_0^q F_0^Q P_l$ $(l = 1, \cdots, A)$ are automorphic forms of the set $L(\Delta, gq + GQ, v^q)$; by (8) and the theorem, they are linearly dependent. Consequently, the automorphic function χ satisfies an algebraic equation of degree q whose coefficients are polynomials in $\chi_1, \cdots, \chi_\nu$ and not all identically 0. Since q is fixed, the automorphic functions with the group Δ form an algebraic field with exactly ν independent elements.

INSTITUTE FOR ADVANCED STUDY

[5] M. Sugawara, *Über eine allgemeine Theorie der Fuchsschen Gruppen und Theta-Reihen*, Ann. of Math. (2) 41, pp. 488–494 (1940).

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

ON THE DERIVATIVES OF THE SECTIONS OF BOUNDED POWER SERIES[1]

BY RHODA MANNING

(Received January 8, 1942)

1. Introduction

Let $f(z)$ represent a power series convergent in the open unit circle $|z| < 1$ and satisfying the condition $|f(z)| \leqq 1$ in $|z| < 1$. It is well known that the sections $s_n(z)$ of $f(z)$ are not in general bounded in the open unit circle $|z| < 1$.[2] In 1925 L. Fejér[3] proved that the sections $s_n(z)$, for all such functions $f(z)$, satisfy the condition $|s_n(z)| \leqq 1$ in the circle $|z| \leqq \frac{1}{2}$ for all n, and that this number $\frac{1}{2}$ cannot in general be replaced by a larger number.

Let r_n denote the radius of the largest circle $|z| \leqq r_n$ in which the sections $s_n(z)$, for all functions $f(z)$ of the above type, satisfy the condition $|s_n(z)| \leqq 1$. I. Schur and G. Szegö,[4] extending Fejér's result, proved that the radii r_n constitute a monotone increasing sequence of algebraic numbers having the limit unity. They also studied the subclass of all functions $f(z)$ satisfying the additional condition $f(0) = 0$, and showed that, for all such functions $f(z)$, the radius R_n of the largest circle $|z| \leqq R_n$ in which the condition $|s'_{n+1}(z)| \leqq 1$ holds, for odd n, $n \geqq 1$, satisfies the algebraic equation

$$1 - 2r - r^2 - (2n + 4)r^{n+1} - (2n + 2)r^{n+2} = 0.$$ [5]

Hence the sequence $\{R_n\}$, n odd, is ever increasing. The object of this note is to discuss the determination of the radii R_n in the case when n is even. The author has found in this case that the R_n, provided $n \geqq 12$, satisfy the similar equation

$$1 - 2r - r^2 + (2n + 4)r^{n+1} + (2n + 2)r^{n+2} = 0.$$

Hence for even n, $n \geqq 12$, the sequence $\{R_n\}$ is ever decreasing. Both sequences have the common limit $\rho = 2^{\frac{1}{2}} - 1$, the only positive root of the equation $1 - 2r - r^2 = 0$.

---

[1] Presented to the Society, December 2, 1939.

[2] L. Fejér, *Über gewisse Potenzreihen an der Konvergenzgrenze*, Sitzungsber. der math.-physik. Klasse der Bayer. Akad. der Wiss., 1910, Nr. 3.

[3] L. Fejér, *Über die Positivität von Summen, die nach trigonometrischen oder Legendreschen Funktionen fortschreiten (Erste Mitteilung)*, Acta litt. ac sci. regiae univ. hung. Francisco-Josephinae, sectio sci. math., vol. 2, 1925, pp. 75–86.

[4] I. Schur and G. Szegö, *Über die Abschnitte einer im Einheitskreise beschränkten Potenzreihe*, Sitzungsber. der Preuss. Akad. der Wiss., physik.-math. Klasse, 1925, in particular pp. 545–555.

[5] Loc. cit. (4), p. 560.

It follows that the sections of $f'(z)$, for even n, $n \geqq 12$, in general remain bounded by unity "longer" than the function $f'(z)$ itself, an unusual occurrence in this type of problem. As another immediate consequence of the theorem, we regain the well known fact that the derivative $f'(z)$ cannot exceed unity in absolute value in the circle $|z| \leqq 2^{\frac{1}{2}} - 1$, and that the bound $2^{\frac{1}{2}} - 1$ cannot in general be replaced by a larger one.[6]

In a thesis submitted to Stanford University[7] the author has shown, by treating each case separately, that the numbers R_n, for the values of n excluded by the theorem, are also algebraic, and that they satisfy the following order relations:

$$R_1 < R_2 < R_3 < R_4 < R_5 < R_6 < R_7 < R_9 < R_8 < R_{11} < R_{13} < \cdots$$
$$\cdots < \rho = 2^{\frac{1}{2}} - 1 < \cdots < R_{14} < R_{12} < R_{10}.$$

To facilitate computation of the radii R_n, $n \geqq 12$, an asymptotic expression for R_n is given, of the form

$$R_n = \rho + (-1)^n a_n \rho^{n+1} + b_n \rho^{2n+1} + (-1)^n c_n' \rho^{3n+1},$$

where

$$a_n = \left(n + 1 + \frac{2^{\frac{1}{2}}}{2}\right), \qquad b_n = (n+1)(n+2)a_n - \tfrac{1}{4}(2 - 2^{\frac{1}{2}})a_n^2,$$

and $0 < c_n' < 2a_n(n+1)^2(n+2)^2$.

Finally, a closely related theorem, stated by I. Schur and G. Szegö for odd values of n, is generalized to include large even values of n.[8]

2. Main theorem

Let $f(z)$ represent a power series convergent in the open unit circle $|z| < 1$ and satisfying the conditions $|f(z)| \leqq 1$ in $|z| < 1$ and $f(0) = 0$. Let R_n denote the radius of the largest circle $|z| \leqq R_n$ in which the section $s_{n+1}'(z)$, for all power series $f(z)$, satisfies the condition $|s_{n+1}'(z)| \leqq 1$. If $n \geqq 1$, $n \neq 2, 4, 6, 8, 10$, n an integer, then the radius R_n is the smallest positive root of the algebraic equation

$$G_n(r) = 1 - 2r - r^2 + (-1)^n[(2n+4)r^{n+1} + (2n+2)r^{n+2}] = 0.$$

Proof. It has been shown[9] that the radius of the largest circle $|z| \leqq R_n$ in which the condition $|s_{n+1}'(z)| \leqq 1$ holds, is the maximum value of r for which the harmonic polynomial

$$T_n(r, \phi) = \tfrac{1}{2} + 2r\cos\phi + 3r^2\cos 2\phi + \cdots + (n+1)r^n \cos n\phi$$

[6] J. Dieudonné, *Polynomes et fontions bornées d'une variable complex*, École Normale Supérieure, Annales Scientifiques, vol. 48, 1930–31, p. 352.

[7] Dissertation, Stanford University, June 1941.

[8] Loc. cit. (4), pp. 558–559.

[9] Loc. cit. (5).

remains non-negative, for all real values of ϕ. Let us denote the product

$$2(1 - 2r\cos\phi + r^2)^2 \cdot T_n(r, \phi) = 1 - 4r^2 + 4r^3\cos\phi - r^4$$
$$- (2n+4)r^{n+1}\cos(n+1)\phi + 2r^{n+2}[(2n+4)\cos n\phi + (n+1)\cos(n+2)\phi]$$
$$- 2r^{n+3}[(n+2)\cos(n-1)\phi + (2n+2)\cos(n+1)\phi] + (2n+2)r^{n+4} \cdot \cos n\phi$$

by $F_n(r, \phi)$. Since the cosine is an even function, we need only consider values of ϕ satisfying $0 \leqq \phi \leqq \pi$.

If n is odd, then $F_n(r, \pi)$ is an obvious lower boundary for $F_n(r, \phi)$. Since further $F_n(r, \pi) = (1 + r)^2 \cdot G_n(r)$, R_n is the only positive root of the equation $G_n(r) = 0$.

Now let n be even. If $r = 0.42$ and $n \geqq 8$, then

$$F_n(r, \pi) = (1 + r)^2(1 - 2r - r^2 + (2n+4)r^{n+1} + (2n+2)r^{n+2})$$
$$< (1 + r)^2(-0.0164 + 0.0113) < 0,$$

whence $R_n < 0.42$, $n \geqq 8$. Hence we may restrict our proof of the inequality $F_n(r, \phi) \geqq F_n(r, \pi)$, $n \geqq 12$, to values of r contained in the interval $0 < r < 0.42$. Now

$$F_n(r, \phi) - F_n(r, \pi) = 4r^3(1 + \cos\phi) - 2r^{n+1}[(n+2)(1 + \cos(n+1)\phi)$$
$$+ \{(2n+4)(1 - \cos n\phi) + (n+1)(1 - \cos(n+2)\phi)\}r$$
$$+ \{(n+2)(1 + \cos(n-1)\phi) + (2n+2)(1 + \cos(n+1)\phi)\}r^2$$
$$+ (n+1)(1 - \cos n\phi)r^3].$$

On dividing this difference by the positive quantity $2r^3(1 + \cos\phi)$, we notice that all the terms except the first in the resulting expression are of the form

$$-c\,\frac{1 + (-1)^{k-1}\cos k\phi}{1 + \cos\phi} = -c\,\frac{1 - \cos k(\pi - \phi)}{1 - \cos(\pi - \phi)},$$

$c > 0$, $k = 1, 2, 3, \cdots$. It is easily verified that this expression is never less than $-ck^2$, and that it attains this value for $\phi = \pi$. Hence the inequality to be proved is equivalent to

$$\lim_{\phi \to \pi} \frac{F_n(r, \phi) - F_n(r, \pi)}{2r^3(1 + \cos\phi)} = 2 - r^{n-2}[(n+2)(n+1)^2$$
$$+ \{(2n+4)n^2 + (n+1)(n+2)^2\}r + \{(n+2)(n-1)^2$$
$$+ (2n+2)(n+1)^2\}r^2 + (n+1)n^2r^3] \geqq 0,$$

which holds since we can satisfy simultaneously

$$r^{n-2}(n+2)^3(1 + r)^3 < 2, \qquad 0 < r < 0.42 \quad \text{and} \quad n \geqq 12.$$

Hence for even n, $n \geqq 12$, and $0 < r < 0.42$,

$$F_n(r, \phi) \geqq F_n(r, \pi) = (1 + r)^2 \cdot G_n(r),$$

whence R_n is the smallest positive root of the algebraic equation $G_n(r) = 0$.

3. Asymptotic Inequalities

It will be assumed in the two cases which follow that $n \geqq 12$.

CASE 1. *n odd.* The following derivation of an upper bound for R_n depends on the remark that R_n is the only positive root of the equation $G_n(r) = 0$. Since $G_n(0) = 1 > 0$ and $G_n(r)$ is ever decreasing for positive values of r, we conclude that if $G_n(r) < 0$ for $r = r_0$, say, then $R_n < r_0$.

We shall suppose that $r = \rho(1 - x)$, where $x = a_n\rho^n - b_n\rho^{2n}$, $(n + 1 + \frac{1}{2}2^{\frac{1}{2}}) = a_n$, $b_n = (n + 1)(n + 2)a_n - \frac{1}{4}(2 - 2^{\frac{1}{2}})a_n^2$, and show that with this choice of r, $G_n(r) < 0$. We note that $0 < x < a_n\rho^n < 0.0005$, since $\rho^{12} = 0.0000255 \cdots$ and $n^3\rho^n$ is a decreasing function of n. Therefore $r > 0$ and for $k = 1, 2$, $r^{n+k} > \rho^{n+k}[1 - (n + k)x]$.[10] Hence

$$\begin{aligned} G_n(r) &< 1 - 2r - r^2 - (2n + 4)\rho^{n+1}[1 - (n + 1)x] - (2n + 2)\rho^{n+2}[1 - (n + 2)x] \\ &= 2(2^{\frac{1}{2}})\rho x - \rho^2x^2 - 2(2^{\frac{1}{2}})\rho^{n+1}[a_n - (n + 1)(n + 2)x] \\ &= 2(2^{\frac{1}{2}})\rho(a_n\rho^n - b_n\rho^{2n}) - \rho^2(a_n^2\rho^{2n} - 2a_nb_n\rho^{3n} + b_n^2\rho^{4n}) \\ &\quad - 2(2^{\frac{1}{2}})\rho^{n+1}[a_n - (n + 1)(n + 2)a_n\rho^n + (n + 1)(n + 2)b_n\rho^{2n}] \\ &= - b_n\rho^{3n+1}[2(2^{\frac{1}{2}})(n + 1)(n + 2) - 2a_n\rho + b_n\rho^{n+1}] < 0. \end{aligned}$$

To derive a lower bound for R_n, we notice that if $G_n(r) > 0$, then $r < R_n$. Set $r = \rho(1 - x)$, where $x = a_n\rho^n - b_n\rho^{2n} + c_n\rho^{3n}$, with a_n, b_n as before, and $c_n = 2a_n(n + 1)^2(n + 2)^2$. On expanding $(1 - x)^{n+k}$ as an exponential series with $(n + k)\log(1 - x)$ as argument, we find $r^{n+k} < \rho^{n+k}[1 - (n + k)x + n^2x^2]$, $k = 1, 2$. Hence

$$\begin{aligned} G_n(r) &> 2(2^{\frac{1}{2}})\rho x - \rho^2x^2 - 2(2^{\frac{1}{2}})\rho^{n+1}[a_n - (n + 1)(n + 2)x + a_nn^2x^2] \\ &> 2(2^{\frac{1}{2}})\rho^{n+1}(a_n - b_n\rho^n + c_n\rho^{2n}) - a_n^2\rho^{2n+2} - 2(2^{\frac{1}{2}})\rho^{n+1}[a_n \\ &\quad - (n + 1)(n + 2)\rho^n(a_n - b_n\rho^n + c_n\rho^{2n}) + n^2a_n^3\rho^{2n}] \\ &= 2(2^{\frac{1}{2}})\rho^{3n+1}[c_n - (n + 1)(n + 2)b_n + (n + 1)(n + 2)c_n\rho^n - n^2a_n^3] > 0, \end{aligned}$$

since $x < a_n\rho^n$ and $c_n > 2(n + 1)(n + 2)b_n > 2n^2a_n^3$.

CASE 2. *n even.* To derive a lower bound for R_n, we notice that $G_n(r) = 0$ has two positive roots, the smaller of which is R_n, and that $G_n(0) = 1 > 0$. Hence if $G_n(r) > 0$ and $G_n'(r) < 0$ simultaneously, then $r < R_n$. Let $r = \rho(1 + x)$, where $x = a_n\rho^n + b_n\rho^{2n}$. A repetition of the argument in the first part of Case 1 gives $r^{n+k} > \rho^{n+k}[1 + (n + k)x]$, $k = 1, 2$, and that $G_n(r) > 0$. That $G_n'(r) < 0$, for $0 < r < \frac{1}{2}$, is trivial.

---

[10] Hardy, Littlewood and Polya, *Inequalities*, p. 40.

To find an upper bound for R_n, we note that if $G_n(r) < 0$, then $R_n < r$. Set $r = \rho(1 + x)$, where $x = a_n\rho^n + b_n\rho^{2n} + c_n\rho^{3n}$. Then $x < (n + 2)\rho^n$, and with a little attention it can be seen that $r^{n+k} = \exp[(n + k) \log \rho(1 + x)] < \rho^{n+k}[1 + (n + k)x + (n + 1)^2x^2]$, $k = 1, 2$. Hence

$$\begin{aligned} G_n(r) &< -2(2^{\frac{1}{2}})\rho x - \rho^2x^2 + 2(2^{\frac{1}{2}})\rho^{n+1}[a_n + (n+1)(n+2)x + a_n(n+1)^2x^2] \\ &< -2(2^{\frac{1}{2}})\rho^{n+1}(a_n + b_n\rho^n + c_n\rho^{2n}) - a_n^2\rho^{2n+2} - 2a_nb_n\rho^{3n+2} \\ &\quad + 2(2^{\frac{1}{2}})\rho^{n+1}[a_n + (n+1)(n+2)\rho^n(a_n + b_n\rho^n + c_n\rho^{2n}) + a_n(n+1)^2 \\ &\qquad\qquad (n+2)^2\rho^{2n}] \\ &= -(2^{\frac{1}{2}})\rho^{3n+1}[c_n - 2b_n(n+1)(n+2) - 2c_n(n+1)(n+2)\rho^n + \\ &\qquad\qquad (2^{\frac{1}{2}})a_nb_n\rho] < 0. \end{aligned}$$

4. A Related Theorem

The method of proof of the main theorem applies to the following theorem: *Let $f(z)$ represent a power series convergent in the open unit circle $|z| < 1$ and satisfying the condition $|f(z)| \leqq 1$ in $|z| < 1$. Let $\alpha < 0, \beta > 0, \alpha + \beta = 1$, and let r_n be the radius of the largest circle $|z| \leqq r_n$ in which the sections $s_n(z)$, for all power series $f(z)$, satisfy the condition $|\alpha s_0(z) + \beta s_n(z)| \leqq 1$. Then for odd n, $n \geqq 1$, and for sufficiently large even n, the radii r_n satisfy the algebraic equation*

$$1 + (\alpha - \beta)r + (-1)^n 2\beta r^{n+1} = 0.$$

Hence $\lim_{n\to\infty} r_n = \dfrac{1}{\beta - \alpha}$.

PROOF. The radius r_n is the maximum value of r for which the cosine polynomial

$$T_n(r, \phi) = \alpha/2 + \beta(\tfrac{1}{2} + r\cos\phi + r^2\cos 2\phi + \cdots + r^n \cos n\phi)$$

remains non-negative, for all real values of ϕ.[11] Let

$$\begin{aligned} F_n(r, \phi) &= 2(1 - 2r\cos\phi + r^2)\cdot T_n(r, \phi) \\ &= \alpha(1 - 2r\cos\phi + r^2) + \beta[1 - r^2 + 2r^{n+2}\cos n\phi - 2r^{n+1}\cos(n+1)\phi]. \end{aligned}$$

If $r \geqq 1$,

$$\begin{aligned} F_n\left(r, \frac{\pi}{n}\right) &\leqq \alpha(1 - r)^2 + \beta\left[1 - r^2 - 2r^{n+1}\left(r + \cos\frac{n+1}{n}\pi\right)\right] \\ &\leqq -2\beta\left(1 + \cos\frac{n+1}{n}\pi\right) < 0. \end{aligned}$$

Hence $r_n < 1$ for all n.

---

[11] Loc. cit. (4), p. 558.

Now

$$F_n(r, \pi) = \alpha(1 + 2r + r^2) + \beta(1 - r^2 + (-1)^n 2r^{n+1} + (-1)^n 2r^{n+2})$$
$$= 1 + 2\alpha r + (\alpha - \beta)r^2 + (-1)^n 2\beta r^{n+1} + (-1)^n 2\beta r^{n+2}$$
$$= (1 + r)(1 + (\alpha - \beta)r + (-1)^n 2\beta r^{n+1}).$$

If n is odd, then $F_n(r, \pi)$ is an obvious lower boundary for $F_n(r, \phi)$, for all real ϕ.

Let n be even. We shall show that the inequality

$$F_n(r, \phi) \geqq F_n(r, \pi), \qquad r = r_n,$$

holds for all real ϕ, and for sufficiently large even n.

Rewritten, it assumes the form

$$-2\alpha r(1 + \cos \phi) - 2\beta r^{n+1}[(1 + \cos (n + 1)\phi + (1 - \cos n\phi)r] \geqq 0.$$

The substitution $\phi = \pi - x$ yields

$$-\alpha(1 - \cos x) - \beta r^n[(1 - \cos (n + 1)x) + (1 - \cos nx)r] \geqq 0,$$

whence, dividing by the positive quantity $(1 - \cos x)$, and taking the limit as $x \to 0$, we obtain the inequality

$$-\alpha \geqq \beta r^n[(n + 1)^2 + n^2 r],$$

which holds for sufficiently large n and $r < 1$. But $r_n < 1$ for all n. Hence r_n is a root of the algebraic equation

$$1 + (\alpha - \beta)r + 2\beta r^{n+1} = 0.$$

Oregon State College,
Corvallis, Ore.

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

THE TRANSFORMATION T OF CONGRUENCES

BY V. G. GROVE

(Received June 11, 1941)

1. Introduction

We propose to study in this paper a certain relationship between congruences in a projective space of three dimensions. Analytical conditions for this relationship, called by us the transformation T, were developed by Cook[1] in a form slightly different from that used in this paper. Fubini[2] also called attention to this relationship somewhat earlier; but neither of these papers showed the relationship between the transformation T and the theory of W-congruences. The present paper is more closely allied with a recent paper by Fubini[3] on W-congruences.

Two congruences Γ and $\bar{\Gamma}$ will be said to be in the *relation of a transformation T*, if the lines of the congruences are in one-to-one correspondence such that

1. corresponding lines are not coplanar,
2. the developables of the congruences correspond, and
3. such that there exists at least three transversal surfaces[4] of each congruence whose tangent planes at their points of intersection with the line of that congruence pass through the corresponding line of the other congruence.

The transformation T is of two types, one of which we have called the asymptotic type, and the other the conjugate type. Associated with each of these types there is a one-parameter family, or pencil of congruences. This pencil seems to be somewhat more general than the pencil[5] defined by Fubini. As in the case of Fubini's pencils, we find that if one congruence of the associated pencil is a W-congruence, all congruences of the pencil are W. Associated with the transformation T are four congruences such that if any three are W-congruences, the other is also.

Let the curves which correspond to the developables of Γ and $\bar{\Gamma}$ be chosen as the parametric curves on the focal surfaces S_z, S_w, $S_{\bar{z}}$, $S_{\bar{w}}$ of Γ and $\bar{\Gamma}$. Then the homogeneous projective coordinates z_i, w_i, $\bar{z}_i$, $\bar{w}_i$, $i = 1, 2, 3, 4$, of the focal points on the lines of the congruences satisfy differential equations of

---

[1] A. J. Cook, *Pairs of rectilinear congruences with generators in one-to-one correspondence*, Trans. Am. Math. Soc., Vol. 32 (1930), pp. 31–46.

[2] G. Fubini, *Su alcune classi di congruenze di rette e sulle transformazione delle Superficie R*, Annali di Matematica, (4), Vol. 1 (1923–24), pp. 241–257.

[3] G. Fubini, *On Bianchi's permutability theorem and the theory of W-congruences*, these Annals, Vol. 41 (1940), pp. 620–638.

[4] A. J. Cook, loc. cit., says that each of the congruence has the intersector property I with respect to the other congruence.

[5] G. Fubini, loc. cit., p. 634.

the form

$$
(1)\qquad
\begin{aligned}
z_u &= f\bar{z} + g\bar{w} + rz + sw, & \bar{z}_u &= \bar{f}z + \bar{g}w + \bar{r}\bar{z} + \bar{s}\bar{w},\\
z_v &= mz + nw, & \bar{z}_v &= \bar{m}\bar{z} + \bar{n}\bar{w},\\
w_u &= Nz + Mw, & \bar{w}_u &= \bar{N}\bar{z} + \bar{M}\bar{w},\\
w_v &= G\bar{z} + F\bar{w} + Sz + Rw, & \bar{w}_v &= \bar{G}z + \bar{F}w + \bar{S}\bar{z} + \bar{R}\bar{w}.
\end{aligned}
$$

The integrability conditions of system (1) may be written in the form

$$
(2a)\qquad
\begin{aligned}
m_u - r_v + nN - sS &= g\bar{G},\\
s_v - n_u + s(R - m) + n(r - M) &= -g\bar{F},\\
f_v + f(\bar{m} - m) + g\bar{S} + sG &= 0,\\
g_v + g(\bar{R} - m) + sF + f\bar{n} &= 0;
\end{aligned}
$$

$$
(2b)\qquad
\begin{aligned}
M_v - R_u + nN - sS &= \bar{g}G,\\
S_u - N_v + S(r - M) + N(R - m) &= -\bar{f}G,\\
F_u + F(\bar{M} - M) + G\bar{s} + gS &= 0,\\
G_u + G(\bar{r} - M) + fS + F\bar{N} &= 0,
\end{aligned}
$$

and two other sets obtained from these by placing bars above the letters where they do not occur and removing those which do occur.

Let us denote by x, y, z, etc. the points whose homogeneous projective coordinates are x_i, y_i, z_i, etc. $(i = 1, 2, 3, 4)$. Let S_x be a transversal surface of $\bar{\Gamma}$ generated by the point x whose coordinates x_i are of the form $x = \bar{w} + \lambda\bar{z}$. It follows that

$$
(3)\qquad
\begin{aligned}
x_u &= (\bar{M} + \lambda\bar{s})x + \lambda(\bar{f}z + \bar{g}w) + L_1\bar{z},\\
x_v &= (\bar{R} + \lambda\bar{n})x + \bar{G}z + \bar{F}w + L_2\bar{z},
\end{aligned}
$$

wherein

$$
\begin{aligned}
L_1 &= \lambda_u - [\bar{s}\lambda^2 + (\bar{M} - \bar{r})\lambda - \bar{N}],\\
L_2 &= \lambda_v - [\bar{n}\lambda^2 + (\bar{R} - \bar{m})\lambda - \bar{S}].
\end{aligned}
$$

The tangent plane to S_x at x passes through the line g of Γ if and only if $L_1 = L_2 = 0$. If one equates the derivatives λ_{uv} and λ_{vu} computed from $L_1 = 0$, and $L_2 = 0$, one finds, by using the integrability conditions (2), that these latter two equations can have analytic solutions only when the equation

$$
(4)\qquad g\bar{F}\lambda^2 - (\bar{g}G + g\bar{G})\lambda + \bar{f}G = 0
$$

is satisfied. It follows from the third property demanded of two congruences that they be in the relation of a transformation T, that the coefficients of (4) vanish. Hence

$$
(5)\qquad g\bar{F} = \bar{g}G + g\bar{G} = \bar{f}G = 0.
$$

In a similar manner one may show that

$$\bar{g}F = \bar{g}G + g\bar{G} = f\bar{G} = 0. \tag{6}$$

Hence the congruences Γ and $\bar{\Gamma}$ are in relation T if and only if conditions (5) and (6) are satisfied.

In general the tangent planes at z and w to S_z, S_w do not coincide. Hence

$$\bar{f}\bar{F} - \bar{g}\bar{G} \neq 0, \qquad fF - gG \neq 0. \tag{7}$$

From (5), (6), (7) we find that the transformation T is of two types; we call these types respectively

(i) the *asymptotic type* if

$$f = \bar{f} = F = \bar{F} = g\bar{G} + \bar{g}G = 0; \tag{8}$$

and the

(ii) *conjugate type* if

$$g = \bar{g} = G = \bar{G} = 0. \tag{9}$$

2. The Asymptotic Type

Let us consider first the asymptotic type of the transformation T. Under the conditions (8), we observe first that the integrability conditions (2) imply that $s = S = \bar{s} = \bar{S} = 0$.

Let λ_1, λ_2 be two distinct solutions of $L_1 = L_2 = 0$, and suppose that these solutions determine the two transversal surfaces S_x, S_y of $\bar{\Gamma}$. Then from (3) we observe that x, y, z and w satisfy equations of the form

$$\begin{aligned} x_u &= ax + bw, & y_u &= a'y + b'w, \\ x_v &= Ax + Bz, & y_v &= A'y + B'z, \\ z_u &= px + qy + rz, & w_u &= Nz + Mw, \\ z_v &= mz + nw, & w_v &= Qx + Py + Rw, \end{aligned} \tag{10}$$

with

$$bp + b'q = 0, \qquad BQ + B'P = 0, \qquad bb'BB' \neq 0. \tag{11}$$

The integrability conditions of system (10) are

$$\begin{aligned} a_v + bQ &= A_u + Bp, & A'_u + B'q &= a'_v + b'P, \\ bP &= Bq, & B'p &= b'Q, \\ B_u + Br &= aB, & b'_v + b'R &= A'b', \\ b_v + bR &= Ab, & B'_u + B'r &= a'B'. \end{aligned} \tag{12}$$

$$\begin{aligned} p_v + Ap &= mp, & P_u + a'P &= MP, \\ q_v + A'q &= mq, & Q_u + aQ &= MQ, \\ r_v + Bp + B'q &= nN + m_u, & R_u + b'P + bQ &= nN + M_v, \\ n_u + Mn &= nr, & N_v + mN &= NR. \end{aligned} \tag{13}$$

From the first of (2a) and (2b) and the third of (13) we find that $g\bar{G} + \bar{g}G = 0$ implies that

$$pB + qB' + b'P + bQ = 0.$$

From (12) and (13) we see that

$$a_v - a'_v = A_u - A'_u,$$
$$r_v + M_v + a_v + a'_v = R_u + m_u + A_u + A'_u,$$

and we may verify that, by a transformation of the form

$$x = \lambda x', \qquad y = \mu y', \qquad z = \nu z', \qquad w = \rho w', \tag{14}$$

we may make

$$\begin{gathered} a = a', \qquad A = A', \\ r + M + 2a = R + m + 2A = 0. \end{gathered} \tag{15}$$

We shall assume that this transformation has been effected. The conditions (15) are maintained by transformations (14) with

$$\lambda/\mu = \text{const.} \qquad \lambda\mu\nu\rho = \text{const.}$$

Again from (12) we note that

$$\frac{\partial}{\partial u} \log \frac{B}{B'} = 0, \qquad \frac{\partial}{\partial v} \log \frac{b}{b'} = 0. \tag{16}$$

And hence from (11) and (16) we may write

$$q = pU, \qquad Q = PV, \qquad b = -b'U, \qquad B' = -BV \tag{17}$$

wherein U and V are respectively functions of u and v alone.

The focal points $\bar{z}$, $\bar{w}$ of $\bar{g}$ are readily found to be determined by the formulas

$$\bar{z} = (B'x - By)/D, \qquad \bar{w} = (by - b'x)/D, \qquad D = bB' - b'B. \tag{18}$$

Hence we can recover equations (1) in the form

$$\begin{aligned} z_u &= pB(1 - UV)\bar{w} + rz, & \bar{z}_u &= w + \bar{r}\bar{z}, \\ z_v &= mz + nw, & \bar{z}_v &= \bar{m}\bar{z} + \bar{n}\bar{w}, \\ w_u &= Nz + Mw, & \bar{w}_u &= \bar{N}\bar{z} + \bar{M}\bar{w}, \\ w_v &= b'P(1 - UV)\bar{z} + Rw, & \bar{w}_v &= \bar{z} + \bar{R}\bar{w}, \end{aligned} \tag{19}$$

wherein

$$\begin{gathered} \bar{r} = 2a - r - (\log D)_u, \qquad \bar{m} = R, \\ \bar{R} = 2A - R - (\log D)_v, \qquad \bar{M} = r, \\ \bar{n} = \frac{BV_v}{b'(1 - UV)}, \qquad \bar{N} = \frac{b'U_u}{B(1 - UV)}, \\ D = b'B(UV - 1), \qquad \Delta = pP(1 - UV). \end{gathered} \tag{20}$$

3. W-congruences in the Asymptotic Case

The differential equations of the asymptotic curves on S_z, S_w are readily found to be respectively

$$pU_u\,du^2 + nP(1 - UV)\,dv^2 = 0, \tag{21}$$
$$Np(1 - UV)\,du^2 + PV_v\,dv^2 = 0.$$

Hence Γ is a W-congruence if and only if the invariant

$$W = nN - \frac{U_u V_v}{(1 - UV)^2}$$

vanishes.

The differential equations of the asymptotic curves on $S_{\bar z}$, $S_{\bar w}$ are found to be

$$\bar N\,du^2 + n\,dv^2 = 0,$$
$$N\,du^2 + \bar n\,dv^2 = 0.$$

It follows that $\bar\Gamma$ is a W-congruence if and only if the invariant

$$\overline{W} = \bar n\bar N - nN$$

vanishes. But from (20) we see that

$$W + \overline{W} = 0. \tag{22}$$

It follows *therefore that, if one of two congruences in the* relation of a transformation T of the asymptotic type is a W-congruence, the other is also.

4. The Transversal Surfaces

If from (10) one eliminates z and w, it will be found that the coordinates of the current points x, and y of S_x, S_y satisfy the equations

$$\begin{aligned} x_{uu} &= \theta x_u - \frac{b'NU}{B}\,x_v + (\ )x, \\ x_{uv} &= Ax_u + ax_v + (\ )x - b'PUy, \\ x_{vv} &= -\frac{Bn}{b'U}\,x_u + \varphi x_v + (\ )x, \end{aligned} \tag{23}$$

$$\begin{aligned} y_{uu} &= \theta' y_u - \frac{b'N}{BV}\,y_v + (\ )y, \\ y_{uv} &= Ay_u + ay_v + (\ )y - b'PVx, \\ y_{vv} &= -\frac{BnV}{b'}\,y_u + \varphi' y_v + (\ )y, \end{aligned} \tag{24}$$

wherein the omitted coefficients are immaterial for our purposes, and wherein

$$\begin{aligned} \theta &= a + M + (\log b'U)_u, \qquad & \theta' &= a + M + (\log b')_u, \\ \varphi &= A + m + (\log B)_v, \qquad & \varphi' &= A + m + (\log BV)_v. \end{aligned} \tag{25}$$

It follows from (23) and (24) that *the curves on S_x, S_y which correspond to the developables of Γ and $\bar{\Gamma}$ form asymptotic nets N_x, N_y on those surfaces. Moreover those surfaces are not ruled surfaces.*

Denote by I_x, I_y the invariants whose vanishing imply that N_x or N_y is isothermally asymptotic. We find that

$$I_x = I_y = \frac{\partial^2}{\partial u\, \partial v} \log \left(\frac{b'^2 N}{B^2 n}\right) = 4(A_u - a_v).$$

But the vanishing of this function, as is seen from the first of (12), implies that $pP - qQ$ vanishes. Hence *neither N_x nor N_y is isothermally asymptotic.*

Suppose S_ξ is a transversal surface of $\bar{\Gamma}$ distinct from S_x, S_y. If the coordinates of ξ are defined by

$$\xi = y + \lambda x$$

it follows that the tangent plane to S_ξ at ξ passes through g if and only if λ is a constant. We may say that the line (ξw) (or (ξ, z)) generates a *pencil of congruences.* They are the asymptotic tangents to the one focal surface S_ξ, the locus of ξ being the line $\bar{g}$. We note that $I_\xi = I_x = I_y$ for every λ.

If we denote the coordinates of the tangent planes to S_x, S_y, $S_{\bar{z}}$, $S_{\bar{w}}$ respectively by ξ, η, ω, ζ we find that these functions satisfy the following system of differential equations

$$\begin{aligned}
&\xi_u = -a\xi + q\omega, &&\eta_u = -a\eta + p\omega,\\
&\xi_v = -A\xi + P\zeta, &&\eta_v = -A\eta + Q\zeta,\\
&\zeta_u = b'\xi + b\eta - M\zeta, &&\omega_u = -N\zeta - r\omega,\\
&\zeta_v = -R\zeta - n\omega, &&\omega_v = B'\xi + B\eta - m\omega.
\end{aligned} \tag{26}$$

We have said[6] that two nets are in relation C if the developables of the congruence of lines joining corresponding points of the nets intersect the sustaining surfaces of the nets in those nets. In particular two conjugate nets in the relation of a fundamental transformation F are in relation C. Two nets in relation C are said to be K_α transforms if

$$\frac{\partial^2}{\partial u\, \partial v} \log \alpha = 0$$

where α is one of the cross ratios of the corresponding points of the nets and the two focal points on the line of the congruence through these points. In particular nets in the relation of a transformation of Koenig are K_α transforms.

We readily verify that $\alpha = bB'/(b'B)$. From (17) we note that $\alpha = UV$. Hence N_x, N_y are K_α transforms in the asymptotic case. Similarly from (26) we see that N_ξ, N_η are also K_α transforms since in that case $\alpha = UV$.

[6] V. G. Grove, *Transformations of Nets*, Trans. Am. Math. Soc., Vol. 30 (1928), pp. 483-497. *Ibid.*, p. 493.

5. The Focal Surfaces

From (19) we see that the functions z and w satisfy equations of the form

$$z_{uv} = mz_u + rz_v + (\ )z,$$
$$w_{uv} = Rw_u + Mw_v + (\ )w,$$

the omitted coefficients being immaterial for our purposes. The nets N_z, N_w have equal point invariants if the respective invariants E_z, E_w defined by

$$E_z = m_u - r_v, \qquad E_w = M_v - R_u, \tag{27}$$

vanish.

It is readily seen that

$$2(E_w - E_z) = I_x = I_y.$$

Hence *not both N_z and N_w can have equal point invariants.*

From (21) we find that N_z and N_w are isothermally conjugate if the respective invariants

$$\begin{aligned} I_z &= \frac{\partial^2}{\partial u\, \partial v} \log \frac{pU_u}{nP(1 - UV)}, \\ I_w &= \frac{\partial^2}{\partial u\, \partial v} \log \frac{PV_v}{Np(1 - UV)}, \end{aligned} \tag{28}$$

vanish. But we may show that

$$I_w - I_z = I_x. \tag{29}$$

Hence *not both N_z, N_w can be isothermally conjugate.*

6. The Conjugate Type

The conjugate type of the transformation T is characterized by the conditions $g = \bar{g} = G = \bar{G} = 0$. Let S_x, S_y be two transversal surfaces of $\bar{\Gamma}$ whose tangent planes pass through the lines g of Γ. Then from (3) and by use of a transformation of the form (14) we may show that x, y, z, w may be made to satisfy equations of the form

$$\begin{aligned} x_u &= bz, & y_u &= b'z, \\ x_v &= Bw, & y_v &= B'w, \\ z_u &= px + qy - Mz + sw, & w_u &= Nz + Mw, \\ z_v &= mz + nw, & w_v &= Qx + Py + Sz - mw, \end{aligned} \tag{30}$$

wherein

$$Bp + B'q = bQ + b'P = 0, \qquad m_u = M_v. \tag{31}$$

The integrability conditions of system (30) are

$$
\begin{array}{llll}
 & b_v + bm = BN, & & B'_u + B'M = b'n, \\
 & B_u + BM = bn, & & b'_v + b'm = B'N, \\
 & P_u + qS = MP, & & p_v + Qs = mp, \\
(32) & Q_u + pS = MQ, & & q_v + Ps = mq, \\
 & sS - nN = 2m_u = 2M_v, & & \\
 & s_v - n_u = 2(ms + Mn), & & S_u - N_v = 2(MS + mN).
\end{array}
$$

System (30) is preserved under all transformations of the form (14) with $\lambda = \text{const.}$, $\mu = \text{const.}$, $\rho\sigma = \text{const.}$

The focal points $\bar{z}$, $\bar{w}$ on the line $\bar{g}$ of $\bar{\Gamma}$ are determined by the formulas

$$
(33) \qquad \bar{z} = (B'x - By)/D, \qquad \bar{w} = (by - b'x)/D, \qquad D = bB' - b'B.
$$

Hence we may recover equations (1) in the form

$$
\begin{array}{llll}
 & z_u = f\bar{z} - Mz + sw, & & \bar{z}_u = z - [M + (\log D)_u]\bar{z} - n\bar{w}, \\
 & z_v = mz + nw, & & \bar{z}_v = m\bar{z} + \bar{n}\bar{w}, \\
(34) & w_u = Nz + Mw, & & \bar{w}_u = \bar{N}\bar{z} + M\bar{w}, \\
 & w_v = F\bar{w} + Sz - mw, & & \bar{w}_v = w - N\bar{z} - [m + (\log D)_v]\bar{w},
\end{array}
$$

wherein

$$
\bar{n} = (BB'_v - B'B_v)/D, \qquad \bar{N} = (b'b_u - bb'_u)/D.
$$

7. W-Congruences in the Conjugate Case

Denote by Γ_{11} the congruence of lines $(x\ z)$, Γ_{22} the congruence of lines (y, w), Γ_{21} that formed by $(y\ z)$, Γ_{12} that formed by $(x\ w)$. Let W_{ij} be an invariant whose vanishing implies that Γ_{ij} is a W-congruence. Four such functions are:

$$
\begin{array}{lll}
 & W_{11} = \dfrac{\partial^2}{\partial u\, \partial v} \log \dfrac{q}{B} - 4m_u, & W_{12} = \dfrac{\partial^2}{\partial u\, \partial v} \log \dfrac{P}{b} - 4m_u, \\
(35) & & \\
 & W_{21} = \dfrac{\partial^2}{\partial u\, \partial v} \log \dfrac{p}{B'} - 4m_u, & W_{22} = \dfrac{\partial^2}{\partial u\, \partial v} \log \dfrac{Q}{b'} - 4m_u.
\end{array}
$$

The congruences Γ, $\bar{\Gamma}$ are W-congruences if the respective invariants

$$
(36) \qquad W = \frac{\partial^2}{\partial u\, \partial v} \log \Delta - 4m_u, \qquad \overline{W} = -\frac{\partial^2}{\partial u\, \partial v} \log D - 4m_u,
$$

$$
(\Delta = pP - qQ)
$$

vanish.

But using (31) we show easily that

$$
(37) \qquad \frac{\Delta}{D} = \frac{qQ}{b'B} = \frac{pP}{bB'}.
$$

Hence

$$W + \overline{W} = W_{11} + W_{22} = W_{12} + W_{21}. \tag{38}$$

From the points x, y, z, w there may be formed three different skew quadrilaterals. From (38) we may say that *if any three of the sides of these quadrilaterals generate W-congruences so also does the fourth side.* If we agree to say that the tangents to a family of asymptotic curves on a surface form a W-congruence, we note that equation (22) is then a special case of (38).

Let S_ξ be a transversal surface of $\overline{\Gamma}$ whose tangent plane at ξ passes through the line g of Γ. Then if

$$\xi = x + \lambda y \tag{39}$$

it follows that $\lambda = \text{const.}$ We find readily that

$$\xi_u = (b + \lambda b')z, \qquad \xi_v = (B + \lambda B')w, \tag{40}$$

and if Γ^λ_{11}, Γ^λ_{12} are the congruences of lines (ξ, z) and (ξ, w) respectively and W^λ_{ij} the corresponding invariants W, then

$$W^\lambda_{11} = W_{11}, \qquad W^\lambda_{12} = W_{12}. \tag{41}$$

We may say that the congruences Γ^λ_{11} and Γ^λ_{12} form pencils. We shall call them *the associated pencils.* It follows from (41) that *if one congruence of an associated pencil is a W-congruence, all congruences of the pencil are W-congruences.*

Denote by ρ_{ij} the focal points (other than x or y) on the lines of the congruences Γ_{ij}. We find that these points are defined by

$$\begin{aligned} \rho_{11} &= Bz - nx, & \rho_{12} &= bw - Nx, \\ \rho_{21} &= B'z - ny, & \rho_{22} &= b'w - Ny. \end{aligned} \tag{42}$$

It may readily be found that

$$\begin{aligned} \rho_{11v} &= (B_v + Bm)z - n_v x, & \rho_{12u} &= (b_u + bM)w - N_u x, \\ \rho_{21v} &= (B'_v + B'm)z - n_v y, & \rho_{22u} &= (b'_u + b'M) - N_u y. \end{aligned}$$

Hence the developables of Γ_{ij} correspond to the developables of Γ and $\overline{\Gamma}$.

Denote by ρ^λ_{ij} the focal points other than ξ (or η) on the lines of the congruences Γ^λ_{ij}. We find that the coordinates of ρ^λ_{ij} are defined by the formulas

$$\begin{aligned} \rho^\lambda_{11} &= (B + \lambda B')z - nx, \\ \rho^\lambda_{12} &= (b + \lambda b')w - Nx. \end{aligned} \tag{43}$$

Hence

$$\rho^0_{11} = \rho_{11}, \qquad \rho^0_{12} = \rho_{12}, \qquad \rho^\infty_{11} = \rho_{21}, \qquad \rho^\infty_{12} = \rho_{22}. \tag{44}$$

It follows from (43) *that each of the focal points of a line of a congruence of a pencil moves along a line as that congruence generates the pencil.* In the pencil as defined by Fubini in the paper cited, one focal point is fixed, the other focal point moves along a line.

8. The Transversal Surfaces

From (30) we may show that the coordinates x, y of the current points of S_x, S_y satisfy the following differential equations

$$\begin{aligned} x_{uu} &= \left(\frac{b_u}{b} - M\right) x_u + \frac{bs}{B} x_v + b(px + qy), \\ x_{uv} &= \left(\frac{b_v}{b} + m\right) x_u + \left(\frac{B_u}{B} + M\right) x_v, \\ x_{vv} &= \frac{BS}{b} x_u + \left(\frac{B_v}{B} - m\right) x_v + B(Qx + Py); \end{aligned} \tag{45}$$

$$\begin{aligned} y_{uu} &= \left(\frac{b'_u}{b'} - M\right) y_u + \frac{b's}{B} y_v + b'(qy + px), \\ y_{uv} &= \left(\frac{b'_v}{b'} + m\right) y_u + \left(\frac{B'_u}{B'} + M\right) y_v, \\ y_{vv} &= \frac{B'S}{b'} y_u + \left(\frac{B'_v}{B'} - m\right) y_v + B'(Py + Qx). \end{aligned} \tag{46}$$

Hence *N_x, N_y are conjugate nets in the relation of a transformation F.*

It follows from (45) and (46) that N_x and N_y have equal point invariants if the respective functions vanish:

$$E_x = \frac{\partial^2}{\partial u\, \partial v} \log \frac{b}{B}, \qquad E_y = \frac{\partial^2}{\partial u\, \partial v} \log \frac{b'}{B'}. \tag{47}$$

Again from (45) and (46) we note that the asymptotic curves on S_x, S_y are given by the respective equations

$$bq du^2 + Bp dv^2 = 0, \qquad b'p du^2 + B'Q dv^2 = 0. \tag{48}$$

But from (31)

$$\frac{bq}{Bp} = \frac{b'p}{B'Q}. \tag{49}$$

Hence *the asymptotic curves on S_x, S_y correspond.*

If we denote the coordinates of the tangent planes to S_x, S_y, $S_{\bar{z}}$, $S_{\bar{w}}$ by ξ, η, ω, ζ respectively, we may write

$$\begin{aligned} &\xi = (x, z, w), \qquad \eta = (y, w, z), \qquad \zeta = (x, y, w) \\ &\omega = (y, x, z). \end{aligned} \tag{50}$$

We find that these functions satisfy the following system of differential equations

$$\begin{aligned} \xi_u &= q\zeta, & \eta_u &= p\zeta, \\ \xi_v &= P\omega, & \eta_v &= Q\omega, \\ \zeta_u &= b'\xi + b\eta + M\zeta - N\omega, & \omega_u &= -s\zeta - M\omega, \\ \zeta_v &= -m\zeta - S\omega, & \omega_v &= B'\xi + B\eta - n\zeta + M\omega. \end{aligned} \tag{51}$$

The equatons of Laplace which ξ, η satisfy are

$$\begin{aligned} \xi_{uv} &= \left(\frac{q_v}{q} - m\right)\xi_u + \left(\frac{P_u}{P} - M\right)\xi_v, \\ \eta_{uv} &= \left(\frac{p_v}{p} - m\right)\eta_u + \left(\frac{Q_u}{Q} - M\right)\eta_v. \end{aligned} \tag{52}$$

It follows therefore that the nets N_x, N_y have equal tangential invariants if the respective invariants

$$E_\xi = \frac{\partial^2}{\partial u\,\partial v}\log\frac{q}{P}, \qquad E_\eta = \frac{\partial^2}{\partial u\,\partial v}\log\frac{p}{Q} \tag{53}$$

vanish. From (47), (49), and (53) we see that

$$E_x + E_\xi = E_y + E_\eta.$$

As is seen from (33) the nets N_x, N_y are in the relation of a transformation K of Koenigs if and only if

$$bB' + b'B = 0. \tag{54}$$

Moreover from (31) and (54) one may show that

$$pP + qQ = 0. \tag{55}$$

But the condition (55) implies that the tangent plans to S_x and S_y at x, y separate the focal planes of the line g of Γ harmonically.

One may show from (32) that the condition (54) implies[7] that $E_x = E_y = 0$, and that (55) implies that $E_\xi = E_\eta = 0$. It follows therefore that if the *two nets* N_x, N_y *are in the relation of a transformation* K *they are also in the relation of a transformation*[7] Ω.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

---

[7] L. P. Eisenhart, *Transformations of Surfaces*, Princeton, 1923, p. 134.

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

ON THE HOMOTOPY GROUPS OF SPHERES AND ROTATION GROUPS[1]

BY GEORGE W. WHITEHEAD

(Received February 11, 1942)

1. Introduction

One of the outstanding problems in modern topology is that of classifying the mappings of an m-dimensional sphere S^m into a topological space X. In terms of the Hurewicz theory of homotopy groups[2] this problem may be phrased as follows: to determine the structure of the m^{th} homotopy group $\pi_m(X)$. Of particular interest is the case where X itself is an n-sphere S^n. In this case the results of Hopf,[3] Freudenthal,[4] and Pontrjagin[5] have led to the solution of the problem for $m \leqq n + 2$. For $m > n + 2$ almost nothing is known concerning the structure of $\pi_m(S^n)$.

That this problem is closely related to the study of homotopy properties of the rotation group R_n of the n-sphere has been shown by Pontrjagin,[5] who has used the one- and two-dimensional homotopy groups of R_n to compute the groups $\pi_{n+i}(S_n)$ $(i = 1, 2)$.

In the present paper we introduce an operation which associates with each mapping $f(S^m \times S^n) \subset S^n$ a mapping $\phi(S^{m+n+1}) \subset S^{n+1}$. This is a generalization of the procedure of Hopf[6] for the case $m = n$. This operation is shown to induce a homomorphism of $\pi_m(R_n)$ into $\pi_{m+n+1}(S^{n+1})$, which for $m = 1, 2$ turns out to be an isomorphism. The connection of this homomorphism with one introduced by Freudenthal[4] is studied.

In a recent paper Freudenthal[7] has announced without proof a very general theorem on extension of mappings, and used this theorem to construct maps of S^{2n-1} on S^n of Hopf invariant 1[6] for all even n. We shall use the above results to construct a counter-example to Freudenthal's theorem. It is further shown that Freudenthal's construction definitely fails if $n > 2$ and $n \equiv 2 \pmod 4$.

2. Preliminary concepts

In Euclidean $(r + 1)$-space $\mathfrak{E}^{r+1}$ let S^r denote the unit sphere, i.e., the set of points $x = (x_1, \cdots, x_{r+1}) \in \mathfrak{E}^{r+1}$ with

$$(1) \qquad |x|^2 = \sum_{i=1}^{r+1} x_i^2 = 1.$$

---

[1] Presented to the American Mathematical Society, December 30, 1941.

[2] W. Hurewicz, Proc. Akad. Amsterdam 38 (1935), pp. 112–119

[3] H Hopf, Math. Ann. 104 (1931), pp. 637-665. We shall refer to this paper as H I.

[4] H. Freudenthal, Comp. Math. 5 (1937), pp. 299–314. We shall refer to this paper as F I.

[5] L. Pontrjagin, C. R. Acad. Sci. URSS 19 (1938), pp. 147–149, 361–363.

[6] H. Hopf, Fund. Math. 25 (1935), pp. 427–440. We shall refer to this paper as H II.

[7] H. Freudenthal, Proc. Akad. Amsterdam 42 (1939), pp. 139–140. We shall refer to this paper as F II.

Let E_i^r $(i = 1, 2)$ be the hemispheres defined by the conditions $x_{r+1} \geqq 0$, $x_{r+1} \leqq 0$, respectively. E^{r+1} denotes the closed $(r+1)$-cell $|x| \leqq 1$ bounded by S^r. We shall refer to the points $x^1 = (0, 0, \cdots, 1)$ and $x^2 = (0, 0, \cdots, -1)$ as the *north* and *south poles*, respectively.

Let Y be a metric space with distance function $\rho(y_1, y_2)$, y^0 a fixed point of Y. By Y^{S^r} we shall mean the space of all mappings[8] $f(S^r) \subset Y$ metrized by

$$\rho(f, g) = \underset{x \in S^r}{\text{L.U.B.}}\, \rho[f(x), g(x)] \qquad (f, g \in Y^{S^r}). \tag{2}$$

Let x^0 be the point of S^r with co-ordinates $(1, 0, \cdots, 0)$. Then $Y^{S^r}(x^0, y^0)$ denotes the subspace of Y^{S^r} consisting of those mappings $f(S^r) \subset Y$ such that $f(x^0) = y^0$. Two mappings $f, g \in Y^{S^r}(x^0, y^0)$ are said to be homotopic if they can be joined by an arc in $Y^{S^r}(x^0, y^0)$. The relation of homotopy is reflexive, symmetric, and transitive and divides the space $Y^{S^r}(x^0, y^0)$ into equivalence classes, called *homotopy classes*. The set of all these homotopy classes we denote by $\pi_r(Y)$. We shall denote the homotopy class of any $f \in Y^{S^r}(x^0, y^0)$ by $\mathbf{f}$.

We define an operation of addition between homotopy classes as follows: let f_i $(i = 1, 2) \in Y^{S^r}(x^0, y^0)$. Let ϕ_i $(i = 1, 2)$ be a mapping of E_i^r on S^r such that (1) $\phi_i(S^{r-1}) = x^0$; (2) $\phi_i(E_i^r - S^{r-1}) \subset S^r$ is a topological map of degree 1. Then we define a mapping $f(S^r) \subset Y$ as follows:

$$f(x) = \begin{cases} f_1[\phi_1(x)] & (x \in E_1^r), \\ f_2[\phi_2(x)] & (x \in E_2^r). \end{cases} \tag{3}$$

It is easily verified that the homotopy class of f depends only on the homotopy classes of f_1 and f_2. Let

$$\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2 . \tag{4}$$

Hurewicz[2] has proved that under the operation of addition so defined the set $\pi_r(Y)$ becomes a group, called the r^{th} *homotopy group* of Y. This group is abelian if $r > 1$; in all the cases we consider here it is also abelian if $r = 1$.

3. The homomorphism H

Let Euclidean $(m + n + 2)$-space be represented as the product space $\mathcal{E}^{m+1} \times \mathcal{E}^{n+1}$, points $x \in \mathcal{E}^{m+n+2}$ being represented by co-ordinates (p, q) $(p \in \mathcal{E}^{m+1}, q \in \mathcal{E}^{n+1})$. Then S^{m+n+1} is defined by

$$|p|^2 + |q|^2 = 1. \tag{5}$$

Let H_1 and H_2 be the subsets of S^{m+n+1} defined by

$$|p| \leqq |q|, \tag{6_1}$$

$$|p| \geqq |q|, \tag{6_2}$$

[8] All mappings are supposed continuous.

respectively. Let

$$(7_1)\qquad \psi_1(p, q) = (p/\mid q\mid, q/\mid q\mid) \qquad ((p, q) \in H_1),$$

$$(7_2)\qquad \psi_2(p, q) = (p/\mid p\mid, q/\mid p\mid) \qquad ((p, q) \in H_2).$$

Evidently $\psi_1 \mid H_1H_2 = \psi_2 \mid H_1H_2$[9] and maps H_1H_2 into $S^m \times S^n$. Denote this mapping by ψ. Then

LEMMA 1. *The mappings ψ_1, ψ_2, and ψ defined above are homeomorphic mappings of H_1 on $E^{m+1} \times S^n$, H_2 on $S^m \times E^{n+1}$, and H_1H_2 on $S^m \times S^n$ respectively.*

Let f be a mapping of $S^m \times S^n$ into S^n. We associate with f the mapping $H(f) = \phi(S^{m+n+1}) \subset S^{n+1}$ as follows: ϕ maps the great circle joining the point $(0, q)$ to the point (p, q) on the great circle joining the north pole z^1 of S^{n+1} to the point $f[\psi^{-1}(p, q)]$, and maps the great circle joining $(p, 0)$ to (p, q) on the great circle joining z^2 to $f[\psi^{-1}(p, q)]$. Evidently $\phi(H_1) \subset E_1^{n+1}$, $\phi(H_2) \subset E_2^{n+1}$, while $\phi = f\psi^{-1}$ on H_1H_2. The functions defining the mapping ϕ are given by

$$(8)\qquad \begin{aligned} \phi_i(p, q) &= 2\mid p\mid\cdot\mid q\mid\cdot f_i(p/\mid p\mid, q/\mid q\mid) && (\mid p\mid\cdot\mid q\mid \neq 0),\\ \phi_i(0, q) &= \phi_i(p, 0) = 0 && (i = 1, \cdots, n+1);\\ \phi_{n+2}(p, q) &= \mid q\mid^2 - \mid p\mid^2. \end{aligned}$$

We use this operation to construct a mapping $\mathbf{H} = \mathbf{H}_{m,n}$ of $\pi_m(R_n)$ into $\pi_{m+n+1}(S^{n+1})$ as follows: let $e \in R_n$ denote the identity mapping of S^n on itself, and let $f \in R_n^{S^m}(p^0, e)$. If $p \in S^m$, $q \in S^n$, let $f^*(p, q)$ denote the point of S^n into which q is carried by the rotation $f(p)$. Let $\phi = H(f^*)$. Then it is easy to verify that $\phi \in S^{n+1^{S^{m+n+1}}}(x^0, z^2)$, where $x^0 = (p^0, 0)$ and z^2 is the south pole of S^{n+1}. Let $\mathbf{H}(\mathbf{f}) = \boldsymbol{\phi}$. Evidently $\mathbf{f} = \mathbf{g}$ implies $\mathbf{H}(\mathbf{f}) = \mathbf{H}(\mathbf{g})$, so that $\mathbf{H}$ is a well-defined mapping of $\pi_m(R_n)$ into $\pi_{m+n+1}(S^{n+1})$. We have further

THEOREM 1. **H** *is a homomorphic mapping of $\pi_m(R_n)$ into $\pi_{m+n+1}(S^{n+1})$.*

For let $\mathbf{f}, \mathbf{g} \in \pi_m(R_n)$, and let h be the constant mapping $h(p) = e$ $(p \in S^m)$. Then $\mathbf{h} = 0$. Hence $\mathbf{f} + \mathbf{h} = \mathbf{f}$, $\mathbf{h} + \mathbf{g} = \mathbf{g}$, so that $\mathbf{H}(\mathbf{f} + \mathbf{h}) = \mathbf{H}(\mathbf{f})$, $\mathbf{H}(\mathbf{h} + \mathbf{g}) = \mathbf{H}(\mathbf{g})$. It is therefore sufficient to prove that

$$(9)\qquad \mathbf{H}(\mathbf{f} + \mathbf{h}) + \mathbf{H}(\mathbf{h} + \mathbf{g}) = \mathbf{H}(\mathbf{f} + \mathbf{g}).$$

Let f', g' be mappings of S^m into R_n defined by

$$(10_1)\qquad f'(p) = \begin{matrix} f[\phi_1(p)] & (p \in E_1^m),\\ h[\phi_2(p)] & (p \in E_2^m); \end{matrix}$$

$$(10_2)\qquad g'(p) = \begin{matrix} h[\phi_1(p)] & (p \in E_1^m),\\ g[\phi_2(p)] & (p \in E_2^m). \end{matrix}$$

Then $\mathbf{f}' = \mathbf{f} + \mathbf{h}$, $\mathbf{g}' = \mathbf{h} + \mathbf{g}$. Let $F = H(f'^*)$, $G = H(g'^*)$.

[9] If $f(x) \subset Y$ and A is a closed subset of X, $f \mid A$ denotes the mapping of A into Y obtained by restricting the range of definition of f to the set A.

Let π_i denote the vertical projection of E_i^{m+n+1} on E^{m+n+1} $(i = 1, 2)$. Then $\pi_i(x) = x$ for $x \in S^{m+n}$. Let $F_0 = F \mid E_1^{m+n+1}$, $H'' = F \mid E_2^{m+n+1}$, $H' = G \mid E_1^{m+n+1}$, $G_0 = G \mid E_2^{m+n+1}$. Then it is easily verified that $H'\pi_1^{-1} = H''\pi_2^{-1}$. Call this mapping H_0. Evidently $F_0(x) = G_0(x) = H_0(x)$ $(x \in S^{m+n})$.

Let H_t $(0 \leqq t \leqq 1)$ be a homotopy of H_0 to x^0 keeping x^0 fixed. Then[10] there exist homotopies F_t, G_t $(0 \leqq t \leqq 1)$ of F_0, G_0 respectively, such that $F_t(x) = G_t(x) = H_t(x)$ $(x \in S^{m+n})$. Let

$$(11_1)\qquad F'_t(x) = \begin{cases} F_t(x) & (x \in E_1^{m+n+1}), \\ H_t[\pi_2(x)] & (x \in E_2^{m+n+1}); \end{cases}$$

$$(11_2)\qquad G'(x) = \begin{cases} H_t[\pi_1(x)] & (x \in E_1^{m+n+1}), \\ G_t(x) & (x \in E_2^{m+n+1}). \end{cases}$$

Evidently $\mathbf{F}'_1 = \mathbf{F}$, $\mathbf{G}'_1 = \mathbf{G}$.

Let

$$(12)\qquad H'_t(x) = \begin{cases} F_t(x) & (x \in E_1^{m+n+1}), \\ G_t(x) & (x \in E_2^{m+n+1}). \end{cases}$$

Then $\mathbf{H}'_0 = \mathbf{H}(\mathbf{f} + \mathbf{g})$, while $\mathbf{H}'_1 = \mathbf{F}'_1 + \mathbf{G}'_1 = \mathbf{F} + \mathbf{G} = \mathbf{H}(\mathbf{f} + \mathbf{h}) + \mathbf{H}(\mathbf{f} + \mathbf{g})$.[11] But $\mathbf{H}'_0 = \mathbf{H}'_1$, which proves the theorem.

4. Relations between the homomorphisms F, G, and H

Let S^{m+n} be the equator of S^{m+n+1}, S^n the equator of S^{n+1}, and let f be a mapping of S^{m+n} into S^n. We associate with the mapping f a mapping $F(f) = \psi(S^{m+n+1}) \subset S^{n+1}$ as follows: ψ maps the great circle joining the north pole x^1 of S^{m+n+1} to the point $x \in S^{m+n}$ on the great circle joining z^1 to $f(x)$, and maps the great circle joining x^2 to x on the great circle joining z^2 to $f(x)$. Evidently $\psi(E_1^{m+n+1}) \subset E_1^{n+1}$, $\psi(E_2^{m+n+1}) \subset E_2^{n+1}$, while $\psi = f$ on S^{m+n}. If $f \in S^{n^{S^{m+n}}}(x^0, y^0)$, then $F(f) \in S^{n+1^{S^{m+n+1}}}(x^0, y^0)$; moreover, f homotopic to g implies $F(f)$ homotopic to $F(g)$. Thus F induces a mapping $\mathbf{F}$ of $\pi_{m+n}(S^n)$ into $\pi_{m+n+1}(S^{n+1})$, which was shown by Freudenthal[4] to be a homomorphism.

Let R_{n-1} be the closed subgroup of R_n consisting of those rotations which leave the north pole fixed. Evidently R_{n-1} is isomorphic with the group of rotations of S^{n-1}. Since $R_{n-1} \subset R_n$, there is a natural homomorphism $\mathbf{G}$ of $\pi_m(R_{n-1})$ into $\pi_m(R_n)$.

THEOREM 2. *The homomorphisms* $\mathbf{F}$, $\mathbf{G}$, *and* $\mathbf{H}$ *are related by*

$$(13)\qquad \mathbf{F}\mathbf{H}_{m,n-1} = \mathbf{H}_{m,n}\mathbf{G}.$$

---

[10] K. Borsuk, Fund. Math. 28 (1937), p. 101.

[11] This follows from the definition of addition in $\pi_{m+n+1}(S^{n+1})$ given by S. Eilenberg (Ann. of Math. 41 (1940), p. 235), which is easily shown to be equivalent to the one given here.

For let $\mathbf{f} \in \pi_m(R_{m-1})$, $g = F[H_{m,n-1}(f)]$, $g' = H_{m,n}[G(f)]$. It is then easily verified that $g = g'$ on S^{m+n}. Moreover $g'(E_1^{m+n+1}) \subset E_1^{n+1}$, $g'(E_2^{m+n+1}) \subset E_2^{n+1}$. Hence for no x is $g'(x) = -g(x)$. It follows that g and g' are homotopic, so that $\mathbf{g} = \mathbf{g}'$.

Let ϕ be a mapping of S^{n-1} into R_{n-1} defined as follows: if $x \in S^{n-1}$, x' is the point in the great circle joining x^1 to x whose angular distance from x^1 is twice that from x^1 to x. Then $\phi(x)$ is that rotation which carries x^1 into x' and leaves each point in the $(n - 2)$-sphere orthogonal to x^1 and x fixed. Let $h = H_{n-1,n-1}(\phi)$. Then it can easily be shown[12] that if n is even h has Hopf invariant 2. We have further:

THEOREM 3. *The kernel of the homomorphism* $\mathbf{F}[\pi_{2n-1}(S^n)] \subset \pi_{2n}(S^{n+1})$ *(n even) is the subgroup of* $\pi_{2n-1}(S^n)$ *generated by* $\mathbf{h}$.

The author has recently shown[13] that $\mathbf{G}(\boldsymbol{\phi}) = 0$; in fact, the kernel of the homomorphism $\mathbf{G}$ is the subgroup of $\pi_{n-1}(R_{n-1})$ generated by $\boldsymbol{\phi}$. It follows from Theorem 2 that $\mathbf{F}[\mathbf{H}_{n-1,n-1}(\boldsymbol{\phi})] = \mathbf{F}(\mathbf{h}) = 0$. Let $\mathbf{g} \in \pi_{2n-1}(S^n)$, and suppose that $\mathbf{F}(\mathbf{g}) = 0$. Then the Hopf invariant of g is even,[14] say $2k$. Let $\mathbf{f} = k\mathbf{h}$. Then $\mathbf{F}(\mathbf{f} - \mathbf{g}) = 0$, and $\mathbf{f} - \mathbf{g}$ has Hopf invariant zero. Hence[15] $\mathbf{f} - \mathbf{g} = 0$, i.e., $\mathbf{g} = \mathbf{f} = k\mathbf{h}$.

THEOREM 4. $\mathbf{H}_{m,n}$ *maps* $\pi_m(R_n)$ *isomorphically for* $m = 1, 2$. $\mathbf{H}_{m,n}$ *maps* $\pi_m(R_n)$ *on* $\pi_{m+n+1}(S^{n+1})$ *for* $m = 1$ *and for* $m = 2$, $n > 1$.

Let $h(S^1) \subset R_1$ be defined by

$$h(x) = \begin{Vmatrix} x_1 & -x_2 \\ x_2 & x_1 \end{Vmatrix}.$$

Then h maps S^1 homeomorphically on R_1, and $\mathbf{h}$ is a generator of the free cyclic group $\pi_1(R_1)$. But $H_{1,1}(h)$ maps S^3 on S^2 with Hopf invariant 1[16] and generates the group $\pi_3(S^2)$. It follows from Theorems 2 and 3 that $\mathbf{H}_{1,n}$ maps $\pi_1(R_n)$ isomorphically on $\pi_{n+2}(S^{n+1})$ for $n > 1$.

Since $\pi_2(R_n) = 0$, it follows that $\mathbf{H}_{2,n}$ is an isomorphism. But $\pi_{n+3}(S^{n+1}) = 0$ for $n > 1$[5], and hence $\mathbf{H}_{2,n}$ maps $\pi_2(R_n)$ on $\pi_{n+3}(S^{n+1})$. This completes the proof of the theorem.

5. Freudenthal's theorem

Freudenthal has recently announced[7] without proof a very general theorem on extension of mappings, and used this theorem to construct maps of S^{2n-1} on S^n with Hopf invariant 1 for all even n.[17] In this section the foregoing results are used to construct a counter-example to Freudenthal's theorem, and to show that the above-mentioned construction fails if $n > 2$ and $n \equiv 2 \pmod 4$.

---

[12] Cf. H II, p. 431.

[13] Ann. of Math. 43 (1942), Theorem 5.

[14] F I, Satz III.

[15] F I, Satz II, 2.

[16] H I, p. 654.

[17] F II, p. 140.

Let points z of Euclidean $2n$-space be represented by complex co-ordinates $(z_1, \cdots, z_n)$. Then S^{2n-1} is represented by the equation $\sum_{i=1}^{n} z_i\bar{z}_i = 1$.

Let P_{n-1} denote complex projective $(n-1)$-space. Then there is a natural mapping $\phi(S^{2n-1}) \subset P_{n-1}$ defined by mapping each point $z \in S^{2n-1}$ into the point of P_{n-1} with the same coordinates. This is evidently a fibre map in the sense of Hurewicz and Steenrod,[18] the fibres being great circles. This mapping $\phi(S^{2n-1}) \subset P_{n-1}$ can be extended to a mapping $\psi(E^{2n}) \subset P_n$, where $\psi(z_1, \cdots, z_n) = (z_1, \cdots, z_n, (1 - \sum z_i\bar{z}_i)^{\frac{1}{2}})$. It is easily verified that ψ is a homeomorphism on $E^{2n} - S^{2n-1}$ and $\psi = \phi$ on S^{2n-1}.

Let X be a topological space, f a mapping of P_{n-1} into X. Then

THEOREM 5. *The mapping $f(P_{n-1}) \subset X$ can be extended to a mapping $f^*(P_n) \subset X$ if and only if the mapping $f\phi(S^{2n-1}) \subset X$ is inessential.*

For if $f\phi$ is inessential, there is a mapping $F(E^{2n}) \subset X$ such that $F = f\phi$ on S^{2n-1}. Let $f^* = F\psi^{-1}$. Then f^* is the required extension. Conversely, if f^* is an extension of f, let $F = f^*\psi$. Then F maps E^{2n} into X and $F = f\phi$ on S^{2n-1}. Hence $f\phi$ is inessential.

Let $g(S^1) \subset R_{2n-1}$ be defined by

$$g(x) = \begin{Vmatrix} x_1 & x_2 & 0 & 0 & \cdots & 0 & 0 \\ -x_2 & x_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & x_1 & x_2 & \cdots & 0 & 0 \\ 0 & 0 & -x_2 & x_1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & x_1 & x_2 \\ 0 & 0 & 0 & 0 & \cdots & -x_2 & x_1 \end{Vmatrix}.$$

Then g is essential or inessential according as n is odd or even. For if $n = 1$, g is a generator of $\pi_1(R_1)$, so that g is essential. If $n = 2$, we have $g(S^1) \subset Q^3$, where Q^3 is the quaternion subgroup of R_3. But $\pi_1(Q^3) = \pi_1(S^3) = 0$. Hence $g = 0$ in $Q^3 \subset R_3$, and g is inessential. The proof is completed by induction.

Let $h = H(g)$. Then it follows from Theorem 4 that $h(S^{2n+1}) \subset S^{2n}$ is essential if n is odd and inessential if n is even. Moreover, it can be directly verfied that there is a mapping $h'(P_n) \subset S^{2n}$ such that $h = h'\phi$, and that h' has degree 1. An application of Theorem 5 gives

THEOREM 6. *If n is even, the mapping $h'(P_n) \subset S^{2n}$ can be extended over P_{n+1}. If n is odd, it cannot be so extended.*

The theorem of Freudenthal's referred to above can be phrased as follows:[19] *Let K be a complex, f a normal mapping[20] of K^q into S^q. Suppose that f can be extended over K^{q+1}. Then f can be extended over K^{2q-1}.*

Let K be a triangulation of P_{n+1}, so that P_n becomes a closed subcomplex L of K. Then $L \subset K^{2n}$. Let h' be the mapping of L into S^{2n} of degree one

[18] W. Hurewicz and N. E. Steenrod, Proc. Nat. Acad. 27 (1941), pp. 60–64.

[19] F II, p. 140.

[20] I.e., $f(K^{q-1}) = x^0$.

described above. Then[21] h' can be deformed into a normal map h''; moreover, h'' can be extended over K if and only if the same is true of h'. Let $H^r(K - L)$ denote the r^{th} cohomology group of $K - L$ with integral coefficients. Then $H^r(K - L) = 0$ for $r < 2n + 2$, while $H^{2n+2}(K - L)$ is a free cyclic group. In particular, $H^{2n+1}(K - L) = 0$. It follows from a theorem of Whitney[22] that h'' can be extended over K^{2n+1}. But h'' cannot be extended over K^{2n+2} for n odd.

Freudenthal's construction of maps of S^{4n-1} on S^{2n} is based on an application of his theorem to the case $K = P_{2n}$, $f(K^{2n}) \subset S^{2n}$, where $f(P_n) \subset S^{2n}$ is of degree one. The argument above shows that this construction breaks down if n is odd and >1; for f cannot even be extended over the subspace P_{n+1} of P_{2n}.

PURDUE UNIVERSITY

---

[21] H. Whitney, Duke Journal 3 (1937), p. 53.
[22] Loc. cit., Theorem 2.

Annals of Mathematics
Vol. 43, No. 4, October, 1942

LINEAR p-ADIC GROUPS AND THEIR LIE ALGEBRAS

By Robert Hooke

(Received December 18, 1941)

1. Introduction

The theory of real or complex Lie groups necessarily treats only those topological groups which are locally connected. It is the object of this paper to develop methods of Lie theory for the opposite case of totally disconnected groups.

We shall consider those groups into which can be introduced local analytic coordinates from a p-adic field K, calling these p-adic Lie groups over K. The concept of the associated Lie algebra over K will be used at once to obtain the usual properties of Lie groups. Bearing in mind the results of Ado[1] on imbedding any Lie algebra of characteristic zero in a Lie algebra of matrices, we shall restrict ourselves to the study of Lie algebras of matrices over K and the p-adic Lie groups contained in the full linear group over K.

Although the coordinates in these groups may be defined only in a certain neighborhood of the identity, all the groups which will occur will be entire groups. It is necessary, however, to identify two subgroups whose intersection is open in each, as is done in the theory of real local Lie groups. It will then be shown in sections 6 and 7 that the usual one-to-one correspondence exists between the subgroups of a Lie group and the subalgebras of its Lie algebra.

The last sections are devoted to certain special groups and their Lie algebras, and in particular to the determination of groups whose Lie algebras are the various "non-exceptional" normal simple Lie algebras over K, which have been classified by Jacobson.

The author wishes to express here his appreciation of the assistance and encouragement of Professor C. Chevalley in the preparation of this paper.

2. Notation and Preliminary Theorems

We shall first list, without proof, a few necessary theorems from p-adic analysis. Unless otherwise noted, these theorems may be found in the papers of Chabauty, [4], and Chevalley, [5].

Let R be the field of rational numbers and p be a fixed prime. There is determined by p a valuation v in R which is defined by

$$v(p) = \rho \qquad 0 < \rho < 1, \rho \text{ a real number.}$$

(For this notation and elementary results, the reader is referred to Albert, [2].) R_p will denote the complete p-adic number field determined from R by the valuation v. This valuation has the properties:

(a) $$v(xy) = v(x)v(y),$$

(b) $$v(x + y) \leqq \max v(x), v(y).$$

[1] See Ado, [1]. The numbers in square brackets will refer to papers in the bibliography.

If x is an element of R_p, then $v(x)$ is defined and is equal to ρ^m where m is some rational integer. If m is not negative, that is, if $v(x) \leqq 1$, then x is called an integer of R_p. The integers of R_p form a ring E_{R_p} all of whose ideals are powers of the prime ideal (p).

Now let K be any finite algebraic extension of R_p. The integers of K are defined as those elements whose irreducible equations over R_p have coefficients which are all integers of R_p. The valuation v has a unique extension to a valuation of K, and the integers of K are those elements k such that $v(k) \leqq 1$.

If we put $d(x, y) = v(x - y)$ in K, then K becomes a metric space which is complete, locally compact, and totally disconnected. The symbol K^n will denote the direct product space of K by itself to n factors and will be called p-adic n-space over K.

A series $\sum a_n$ with terms in K converges if and only if $\lim_{n\to\infty} v(a_n) = 0$. An analytic function defined on K^n is by definition the sum in its region of convergence of a power series of the type

$$\sum a_{h_1h_2\cdots h_n} x_1^{h_1} x_2^{h_2} \cdots x_n^{h_n}$$

which converges for all points $(x_1, x_2, \cdots, x_n)$ in some neighborhood of the origin in K^n. Such a power series can be differentiated term by term to give a new series convergent in the same region. It has the properties of the absolutely convergent series of complex analysis.

We conclude this summary with a list of theorems from p-adic analysis. These will be referred to throughout the paper by the numbers given them.

(1) *Every integer of R_p is a limit of rational integers.*

(2) *Let $f(x_1, x_2, \cdots, x_n)$ be an analytic function defined in a neighborhood D of the origin in K^n. Let $f_i(y_1, y_2, \cdots, y_m)$, $(i = 1, 2, \cdots, n)$ be n analytic functions defined in a neighborhood D' of the origin in K^m such that $f_i(0, 0, \cdots, 0) = 0$, $(i = 1, 2, \cdots, n)$. Then if in f we substitute for each x_i the series f_i, we get a series $f'(y_1, y_2, \cdots, y_m)$ which converges for all points y in D' which are such that the point with coordinates $f_i(y_1, y_2, \cdots, y_m)$ is in D.*

(3) *Let $f_1, f_2, \cdots, f_h$ be h functions of $h + m$ variables $u_1, u_2, \cdots, u_h$; $x_1, x_2, \cdots, x_m$ such that*

$$f_i(0, 0, \cdots, 0) = 0, \qquad i = 1, 2, \cdots, h,$$

and such that

$$\frac{D(f_1, f_2, \cdots, f_h)}{D(u_1, u_2, \cdots, u_h)} \neq 0$$

when the u_i and the x_i are all zero. Then the equations

$$f_i(u_1, u_2, \cdots, u_h; x_1, x_2, \cdots, x_m) = 0, \quad i = 1, 2, \cdots, h,$$

define the u_i as analytic functions of the x_i,

$$u_i = F_i(x_1, x_2, \cdots, x_m), \qquad i = 1, 2, \cdots, h,$$

where

$$F_i(0, 0, \cdots, 0) = 0, \qquad i = 1, 2, \cdots, h.$$

(4) (Cf. Lutz, [14].) *A system of differential equations*

$$dy_i/dt = f_i(t, y_1, y_2, \cdots, y_n), \qquad i = 1, 2, \cdots, n,$$

where the f_i are analytic functions, has, in some neighborhood of $t = y_i = 0$, one and only one solution in the form

$$y_i = g_i(t), \qquad i = 1, 2, \cdots, n,$$

where the g_i are analytic functions with the initial conditions

$$g_i(0) = 0, \qquad i = 1, 2, \cdots, n.$$

(5) *If x is an element of K and $v(x) < \rho^{1/(p-1)}$, the series*

$$\exp x = 1 + x + x^2/2! + \cdots + x^n/n! + \cdots$$

converges, and $v((\exp x) - 1) = v(x)$. We have, as in ordinary analysis, $\exp (x + y) = (\exp x)(\exp y)$, when all of these exist.

(6) *If x is an element of K, the series*

$$\log x = (x - 1) - (x - 1)^2/2 + \cdots + (-1)^{n-1}(x - 1)^n/n + \cdots$$

converges when $v(x - 1) < 1$.

(7) *If $v(x) < \rho^{1/(p-1)}$, $\log (\exp x)$ exists and is equal to x. If $v(x - 1) < \rho^{1/(p-1)}$, $\exp (\log x)$ exists and is equal to x.* To prove the second statement, we need only show that $v(\log x) < \rho^{1/(p-1)}$. We have

$$v(\log x) \leqq \max (v_n), \quad \text{where} \quad v_n = v[(x - 1)^n/n].$$

If $v(n) = \rho^{\alpha}$, then $n \geqq p^{\alpha}$, and we have
$\alpha \leqq p^{\alpha-1} + p^{\alpha-2} + \cdots + p + 1 = (p^{\alpha} - 1)/(p - 1) \leqq (n - 1)/(p - 1)$.
Hence $v_n < (\rho^{1/(p-1)})^n/(\rho^{(n-1)/(p-1)}) \leqq \rho^{1/(p-1)}$. Q.E.D.

(8) *If an analytic function $f(t)$ is equal to zero for a sequence of values of t approaching zero and $f(0) = 0$, then $f(t)$ is identically zero.* (The proof is as in ordinary analysis.)

3. Matrices in a p-adic Field

Given a field K, we shall denote by K_n the full matric algebra of n-rowed square matrices over K, and by K_{nl} the Lie algebra obtained from K_n by defining the commutator by

$$[x, y] = xy - yx.$$

This will be called a pure commutator of degree 2 in x and y. If c^m is a pure commutator of degree m in x and y, then by definition, $[c^m, x]$ and $[c^m, y]$ are pure commutators of degree $m + 1$ in x and y. The full linear group of order n over K we shall denote simply by G_K, since n will be fixed throughout.

A sequence of matrices is said to converge, if for each fixed pair i, j, the elements in the i^{th} row and j^{th} column of these matrices form a convergent sequence. The limit matrix is the matrix whose elements are the limits of these sequences.

Given a matrix $A = \| a_{ij} \|$ in K_n, there can be defined a weak type of valuation V in K_n by putting $V(A) = \max v(a_{ij})$. This valuation satisfies the conditions

$$V(A + B) \leqq \max V(A), V(B), \tag{a'}$$

$$V(tA) = v(t)V(A), \qquad t \text{ in } K, \tag{b'}$$

$$V(AB) \leqq V(A)V(B). \tag{c'}$$

Now a sequence of matrices $A_m = \| a_{ij}^{(m)} \|$ with elements in a p-adic field K converges if and only if

$$\lim_{m\to\infty} v(a_{ij}^{(m)} - a_{ij}^{(m+1)}) = 0 \text{ for all } i, j.$$

We have, however,

$$v(a_{ij}^{(m)} - a_{ij}^{(m+1)}) \leqq V(A_m - A_{m+1})$$

for each i and j, and there exists for each m a pair of integers i and j such that the equality sign holds. We thus get the same condition for the convergence of a sequence of matrices as for a sequence of elements in K, using now the valuation V. It can be seen, therefore, that in dealing with sequences and series of matrices, we get automatically the same theorems for commutative matrices with the valuation V that are proved for elements in K with the valuation v, provided these theorems are proved using only the fact that the valuation v satisfies the conditions (a′), (b′), (c′).

In particular, we have the following theorems.

THEOREM 1. *If X is in K_n and $V(X) < \rho^{1/(p-1)}$, the series*

$$\exp tX = I + tX + \cdots + t^n X^n/n! + \cdots \quad (I \textit{ the identity matrix})$$

converges for $v(t) \leqq 1$ and we have $V(\exp tX - 1) = V(tX)$. Also for any matrix Y there exists a real number $r > 0$ such that $\exp tY$ converges for $v(t) \leqq r$.

THEOREM 2. *If $V(A - I) < 1$, the series*

$$\log A = (A - I) - (A - I)^2/2 + \cdots + (-1)^{n-1}(A - I)^n/n + \cdots$$

converges. If $V(X) < \rho^{1/(p-1)}$, then $\log(\exp X)$ is defined and equal to X. Also if $V(A - I) < \rho^{1/(p-1)}$, then $\exp(\log A)$ is defined and equal to A.

If X and Y are commutative matrices, it can be shown in the usual way that $\exp(X + Y) = (\exp X)(\exp Y)$ when these exist. We shall need the following generalization of this fact for non-commutative matrices:

THEOREM 3. *Let X and Y be matrices in K_n such that*

$$V(X), V(Y) < \rho^{1/(p-1)}.$$

i) *There is a matrix Z defined by $\exp Z = (\exp X)(\exp Y)$.*

ii) $Z = X + Y + \lim_{m\to\infty} f_m$, *where the* f_m *are linear combinations of higher commutators of* X *and* Y *with rational coefficients.*

PROOF. i) By Theorem 1, $V((\exp X) - I) = V(X)$ and $V((\exp Y) - I) = V(Y)$. Hence

$$\begin{aligned} V[(\exp X)(\exp Y) - I] &= V[(\exp X - I)(\exp Y - I) \\ &\quad + (\exp X - I) + (\exp Y - I)] \\ &\leqq \max V(X), V(Y). \end{aligned}$$

By Theorem 2, therefore, $Z = \log[(\exp X)(\exp Y)]$ exists and $\exp Z = (\exp X)(\exp Y)$.

ii) If X and Y are real matrices, it has been shown by Hausdorff, [7], that

$$Z = X + Y + \sum_{r=2}^{\infty} \sum_{s=1}^{\alpha_r} a^{rs} f^{rs}(X, Y),$$

where f^{rs} is a pure commutator of degree r in X and Y and each a^{rs} is a rational number. Each α_r is finite.

It follows that

$$Z = X + Y + \lim_{m\to\infty} \sum_{r=2}^{m} \sum_{s=1}^{\alpha_r} a^{rs} f^{rs}(X, Y).$$

Each f^{rs} can be expressed as a polynomial which is homogeneous of degree r in X and Y. Let us now put

$$F^r(X, Y) = \sum_{s=1}^{\alpha_r} a^{rs} f^{rs}(X, Y).$$

Then $Z = X + Y + \lim_{m\to\infty} \sum_{r=2}^{m} F^r(X, Y)$, where each F^r is a homogeneous polynomial of degree r in X and Y.

Expressing this series as n^2 series in the elements of the matrices therein, we have

$$Z_{ij} = X_{ij} + Y_{ij} + \lim_{m\to\infty} \sum_{r=2}^{m} F_{ij}^r(X_{11}, \cdots, X_{nn}; Y_{11}, \cdots, Y_{nn}),$$
$$i, j = 1, 2, \cdots, n.$$

It is also true, however, that for sufficiently small values of the elements, the series

$$Z_{ij} = \{\log[(\exp X)(\exp Y)]\}_{ij}, \qquad i, j = 1, 2, \cdots, n,$$

converge. These may be written

$$Z_{ij} = X_{ij} + Y_{ij} + \lim_{m\to\infty} \sum_{r=2}^{m} P_{ij}^r(X_{11}, \cdots, X_{nn}; Y_{11}, \cdots, Y_{nn})$$
$$i, j = 1, 2, \cdots, n$$

A sequence of matrices is said to converge, if for each fixed pair i, j, the elements in the i^{th} row and j^{th} column of these matrices form a convergent sequence. The limit matrix is the matrix whose elements are the limits of these sequences.

Given a matrix $A = \| a_{ij} \|$ in K_n, there can be defined a weak type of valuation V in K_n by putting $V(A) = \max v(a_{ij})$. This valuation satisfies the conditions

$$V(A + B) \leqq \max V(A), V(B), \tag{a'}$$

$$V(tA) = v(t)V(A), \qquad t \text{ in } K, \tag{b'}$$

$$V(AB) \leqq V(A)V(B). \tag{c'}$$

Now a sequence of matrices $A_m = \| a_{ij}^{(m)} \|$ with elements in a p-adic field K converges if and only if

$$\lim_{m \to \infty} v(a_{ij}^{(m)} - a_{ij}^{(m+1)}) = 0 \text{ for all } i, j.$$

We have, however,

$$v(a_{ij}^{(m)} - a_{ij}^{(m+1)}) \leqq V(A_m - A_{m+1})$$

for each i and j, and there exists for each m a pair of integers i and j such that the equality sign holds. We thus get the same condition for the convergence of a sequence of matrices as for a sequence of elements in K, using now the valuation V. It can be seen, therefore, that in dealing with sequences and series of matrices, we get automatically the same theorems for commutative matrices with the valuation V that are proved for elements in K with the valuation v, provided these theorems are proved using only the fact that the valuation v satisfies the conditions (a′), (b′), (c′).

In particular, we have the following theorems.

THEOREM 1. *If X is in K_n and $V(X) < \rho^{1/(p-1)}$, the series*

$$\exp tX = I + tX + \cdots + t^n X^n/n! + \cdots \quad (I \text{ the identity matrix})$$

converges for $v(t) \leqq 1$ and we have $V(\exp tX - 1) = V(tX)$. *Also for any matrix Y there exists a real number $r > 0$ such that* $\exp tY$ *converges for $v(t) \leqq r$.*

THEOREM 2. *If $V(A - I) < 1$, the series*

$$\log A = (A - I) - (A - I)^2/2 + \cdots + (-1)^{n-1}(A - I)^n/n + \cdots$$

converges. If $V(X) < \rho^{1/(p-1)}$, then $\log(\exp X)$ *is defined and equal to X. Also if $V(A - I) < \rho^{1/(p-1)}$, then* $\exp(\log A)$ *is defined and equal to A.*

If X and Y are commutative matrices, it can be shown in the usual way that $\exp(X + Y) = (\exp X)(\exp Y)$ when these exist. We shall need the following generalization of this fact for non-commutative matrices:

THEOREM 3. *Let X and Y be matrices in K_n such that*

$$V(X), V(Y) < \rho^{1/(p-1)}.$$

i) *There is a matrix Z defined by* $\exp Z = (\exp X)(\exp Y)$.

ii) $Z = X + Y + \lim_{m\to\infty} f_m$, *where the* f_m *are linear combinations of higher commutators of* X *and* Y *with rational coefficients.*

PROOF. i) By Theorem 1, $V((\exp X) - I) = V(X)$ and $V((\exp Y) - I) = V(Y)$. Hence

$$\begin{aligned} V[(\exp X)(\exp Y) - I] &= V[(\exp X - I)(\exp Y - I) \\ &\qquad + (\exp X - I) + (\exp Y - I)] \\ &\leqq \max V(X), V(Y). \end{aligned}$$

By Theorem 2, therefore, $Z = \log[(\exp X)(\exp Y)]$ exists and $\exp Z = (\exp X)(\exp Y)$.

ii) If X and Y are real matrices, it has been shown by Hausdorff, [7], that

$$Z = X + Y + \sum_{r=2}^{\infty} \sum_{s=1}^{\alpha_r} a^{rs} f^{rs}(X, Y),$$

where f^{rs} is a pure commutator of degree r in X and Y and each a^{rs} is a rational number. Each α_r is finite.

It follows that

$$Z = X + Y + \lim_{m\to\infty} \sum_{r=2}^{m} \sum_{s=1}^{\alpha_r} a^{rs} f^{rs}(X, Y).$$

Each f^{rs} can be expressed as a polynomial which is homogeneous of degree r in X and Y. Let us now put

$$F^r(X, Y) = \sum_{s=1}^{\alpha_r} a^{rs} f^{rs}(X, Y).$$

Then $Z = X + Y + \lim_{m\to\infty} \sum_{r=2}^{m} F^r(X, Y)$, where each F^r is a homogeneous polynomial of degree r in X and Y.

Expressing this series as n^2 series in the elements of the matrices therein, we have

$$Z_{ij} = X_{ij} + Y_{ij} + \lim_{m\to\infty} \sum_{r=2}^{m} F_{ij}^r(X_{11}, \cdots, X_{nn}; Y_{11}, \cdots, Y_{nn}), \qquad i, j = 1, 2, \cdots, n.$$

It is also true, however, that for sufficiently small values of the elements, the series

$$Z_{ij} = \{\log[(\exp X)(\exp Y)]\}_{ij}, \qquad i, j = 1, 2, \cdots, n,$$

converge. These may be written

$$Z_{ij} = X_{ij} + Y_{ij} + \lim_{m\to\infty} \sum_{r=2}^{m} P_{ij}^r(X_{11}, \cdots, X_{nn}; Y_{11}, \cdots, Y_{nn}) \qquad i, j = 1, 2, \cdots, n$$

where P^r is the sum of terms of degree r in the expansion. We now have the equivalent of two power series expansions for Z_{ij} in terms of the n^2 real variables X_{ij}, Y_{ij}. The two series must therefore be identical term by term.

If now X and Y are p-adic matrices, it follows from i) that

$$Z_{ij} = X_{ij} + Y_{ij} + \lim_{m\to\infty} \sum_{r=2}^{m} P^r_{ij}, \qquad i, j = 1, 2, \cdots, n,$$

and so

$$Z_{ij} = X_{ij} + Y_{ij} + \lim_{m\to\infty} \sum_{r=2}^{m} F^r_{ij}, \qquad i, j, = 1, 2, \cdots, n.$$

Returning to the matric expressions, we have

$$\begin{aligned} Z &= X + Y + \lim_{m\to\infty} \sum_{r=2}^{m} F^r(X, Y) \\ &= X + Y + \lim_{m\to\infty} \sum_{r=2}^{m} \sum_{s-1}^{\alpha_r} a^{rs} f^{rs}(X, Y), \end{aligned}$$

since for any finite value of m these are identical. Q.E.D.

4. p-adic Lie Groups

We make the following definition as in the theory of real and complex Lie groups:

Definition. *A p-adic Lie group G is a topological group equipped with a homeomorphic mapping of a neighborhood N of its identity element onto a neighborhood of the origin of K^m which satisfies the condition: If X, Y, Z are in N, $XY = Z$, and X is mapped on $(x_1, x_2, \cdots, x_m)$ in K^m, etc., then*

$$z_i = f_i(x_1, x_2, \cdots, x_m; y_1, y_2, \cdots, y_m), \quad i = 1, 2, \cdots, m,$$

where the f_i are analytic functions. G is then said to have local analytic coordinates.

The group G_K can be shown to be a p-adic Lie group. We introduce a metric topology by defining the distance between two elements A and B as $V(A - B)$. Then if we put each matrix A in the form $A = I + \| a_{ij} \|$, we can define the n^2 p-adic numbers a_{ij} as the coordinates of A. There clearly exists a real number $q > 0$ such that for any set of a_{ij} in K satisfying the inequality $v(a_{ij}) \leqq q$, the matrix whose coordinates are these numbers is non-singular and in G_K. It follows that there exists a neighborhood of I in G_K homeomorphic to a neighborhood of the origin in K^n. Since multiplication is a polynomial operation in this group, these coordinates are analytic. Since K is locally compact, so is G_K.

A complete system of neighborhoods of I is furnished by the set of spheres S_r consisting of those matrices whose coordinates satisfy for some fixed real number r the inequality

$$v(a_{ij}) \leqq r.$$

Since v is a discrete valuation, these spheres are both open and closed.

For $r < 1$, S_r is always a subgroup. From the nature of the valuation, the product of two elements in S_r is again in S_r. The existence of inverses follows from the fact that the determinant of any matrix in S_r has value 1. These spheres form an infinite descending chain of open and closed subgroups whose intersection is I, so G_K is totally disconnected. It is the existence of these open subgroups which creates most of the difference between the theory of p-adic Lie groups and that of real Lie groups.

5. The Lie Algebras of G_K and its Subgroups

The Lie algebra (infinitesimal group) of a p-adic Lie group may be defined exactly as in the case of real Lie groups (cf. Pontrjagin, [15]). Many relations here may be proved exactly as in the ordinary case, and so their proofs will be omitted. We shall depart from the usual procedure, however, by using the Lie algebra to obtain conditions for the subgroups of a Lie group to be themselves Lie groups.

An *analytic curve* in G_K is an analytic function

$$F(t) = I + X_1 t + X_2 t^2 + \cdots$$

where the X_i are matrices in K_n. The series converges to a non-singular matrix, that is, to an element of G_K, for all t such that $v(t)$ is less than some fixed real number q. The expressions $f_i(t)$ denote the coordinates of $f(t)$, and the *tangent* (at I) of $f(t)$ is the matrix X_1, as usual.

$f(t)$ is a *one-parameter-subgroup* if $f(t_1 + t_2) = f(t_1)f(t_2)$. The analytic curve $\exp tX$ for any matrix X is such a subgroup. The one-parameter subgroups here differ from those of ordinary Lie theory in the following respect: If $q > 0$ is any real number, then those values of t for which $v(t) \leqq q$ and for which $f(t)$ is defined give values of $f(t)$ which form an entire group, not merely a local group. The fact that products exist in this group follows from the fact that v satisfies the condition (b).

From Theorems 1 and 2 it is seen that there exists a neighborhood of I in G_K in which for every element A, the matrix $X_A = \log A$ is defined and $\exp tX_A$ is a one-parameter subgroup which is defined for $v(t) \leqq 1$ and passes through A for $t = 1$. All functions of the form $\exp tX$ are one-parameter subgroups in their region of convergence. A converse to this statement is contained in the following theorem.

THEOREM 4. *In G_K there exists a neighborhood of the identity in which there lies one and only one one-parameter subgroup with a given tangent.*

Using the uniqueness theorem for differential equations, (4), this can be proved exactly as in the theory of real Lie groups, (Pontrjagin, [15], pp. 185–187.) It follows that there exists a sphere S about I in which the only one-parameter subgroups are the exponential functions. Given any subgroup H of G_K and any sphere S_r contained in S we shall denote by H_r the subgroup $H \cap S_r$.

DEFINITION. *The Lie algebra L of a subgroup H of G_K is the set of tangents to analytic curves lying in some H_r. This is clearly independent of the choice of the number r.*

The commutator of two elements X, Y in the Lie algebra of G_K is defined as in Pontrjagin, [15], p. 238, and it can be shown that $[X, Y] = XY - YX$. The Lie algebra of G_K is then the Lie algebra K_{nl}, since any X in K_n is the tangent to an analytic curve $\exp tX$ in G_K. From the definition it follows as usual that the set L associated with a subgroup of G_K is a subalgebra of K_{nl}; it also follows from the definition and Theorem 3 that it is an ideal (invariant subalgebra) if and only if H_r is an invariant subgroup.

6. Analytic Subgroups and Subalgebras

DEFINITION. *Two subgroups H' and H'' of G_K are* equivalent *if there exists a sphere S_r around I such that $H'_r = H''_r$.*

It can be shown that this is the same identification that is used in the theory of real local Lie groups, where H' and H'' are identified if their intersection is open in each. It is the object of this section to use this equivalence relation in obtaining a one-to-one correspondence between classes of equivalent subgroups of G_K and subalgebras of K_{nl}.

It has been shown that every subgroup of G_K has a Lie algebra which is a subalgebra of K_{nl}. Conversely, it will now be shown that if L is any subalgebra of K_{nl}, there is a closed subgroup of G_K whose Lie algebra is L. Let H be the totality of elements of G_K of the form $\exp X$, where X is an element of L, and such that $V(\exp X - I) < \rho^{1/(p-1)}$. By Theorem 1 we must have $V(X) < \rho^{1/(p-1)}$. If X and Y are in L and $V(X)$, $V(Y) < \rho^{1/(p-1)}$, we have from Theorem 3 that $(\exp X)(\exp Y) = \exp Z$; here $V(\exp Z - I) < \rho^{1/(p-1)}$ and Z is in L because L is locally compact and Z is a limit of elements of L. H is therefore closed under multiplication. H must clearly contain the identity and the inverses of all its elements, since $\exp 0 = 1$ and $\exp(-X)$ is the inverse of $\exp X$. Hence H is a group. H is closed since it is a homeomorphic map of a bounded and closed subset of the locally compact space L.

DEFINITION. *A subgroup H of G_K is* analytic *if it is equivalent to a subgroup constructed from a subalgebra as above.*

THEOREM 5. *If H is an analytic subgroup of G_K and $f(t)$ is any analytic curve in H with tangent a_1, there exists in H a one-parameter subgroup with tangent a_1.*

PROOF. Let L be the subalgebra from which H has been constructed, and let S_r be a sphere contained in S. Let

$$f(t) = I + a_1t + a_2t^2 + \cdots$$

be the given analytic curve. Let $f(t_1)$ be some point on this curve in H_r. We can then define

$$X_1 = \log f(t_1)$$

and the one-parameter subgroup

$$g_1(u) = \exp uX_1/t_1$$

passes through $f(t_1)$ for $u = t_1$ and lies in H_r for all values of u such that $V[g_1(u) - I] \leqq r$. The matrix X_1 is in L since the group H is defined by L and $\exp X_1$ is in H.

The tangent to the curve $g_1(u)$ is

$$\begin{aligned} X_1' &= X_1/t_1 = (1/t_1) \log f(t_1) \\ &= (1/t_1)[(a_1t_1 + a_2t_1^2 + \cdots) - (a_1t_1 + a_2t_1^2 + \cdots)^2/2 + \cdots] \\ &= a_1 + t_1\epsilon \end{aligned}$$

where ϵ approaches zero with t_1. Let us now consider a sequence of parameters $t_1, t_2, \cdots$ which approach zero. The corresponding one-parameter subgroups $g_1(u), g_2(u), \cdots$ have tangents $X_1' = a_1 + t_1\epsilon$, $X_2' = a_1 + t_2\epsilon, \cdots$

H is closed, and so for any fixed value of u, $\lim_{n\to\infty} g_n(u)$ is a point in H. The set of these points for all values of u under consideration can be shown to form a one-parameter subgroup $g(u)$ which is clearly not merely the identity. From Theorem 4, $g(u)$ must be an exponential function, so it is clear that its tangent must be the limit of the tangents X_n', which is a_1. The element a_1 must be in L, since L is locally compact. Q.E.D.

It follows from this theorem that if H is an analytic subgroup defined from a subalgebra L, then the Lie algebra of H is L. We thus obtain a one-to-one correspondence between the subalgebras of K_{nl} and the classes of equivalent analytic subgroups of G_K.

THEOREM 6. *Let H be an analytic subgroup of G_K with Lie algebra L of order m. There exists in G_K a system D of local analytic coordinates a_i'' in a neighborhood S_r of I such that an element A of G_K is in H_r if and only if*

$$a_i'' = 0, \qquad i = m + 1, \cdots, n^2.$$

The system D arises from the original system by an analytic transformation, and so by (2), *we may say that H is defined by analytic functions.*

PROOF. Let X be any matrix in K_{nl} such that $\exp X$ exists in G_K. We define the functions

$$h_i(X) = h_i(x_1, x_2, \cdots, x_{n^2}) = (\exp X)_i, \qquad i = 1, 2, \cdots, n^2,$$

where the x_i are the elements of the matrix X in K_n and the $(\exp X)_i$ are the coordinates of this element in G_K. We have then $h_i(0, 0, \cdots, 0) = 0$ and $(\partial h_i/\partial x_j)_0 = \delta_j^i$, so the Jacobian of these functions is not zero at the origin. It follows from (3) that the equations

$$a_i = h_i(a_1', a_2', \cdots, a_{n^2}'), \qquad i = 1, 2, \cdots, n^2,$$

can be solved for a_j' in terms of the original coordinates $a_1, a_2, \cdots, a_{n^2}$. By (2), therefore, the numbers a_j' furnish a new local analytic coordinate system in G_K.

What we have done is to assign to each element A of G_K near I the elements of $\log A$ as its coordinates. Now given a subalgebra L of K_{nl}, we can change

the basis of K_{nl} so that a matrix of K_{nl} is in L if and only if its last $n^2 - m$ coordinates are all zero. Such a transformation is algebraic, and analytic, and has a non-vanishing Jacobian at the origin. Relative to this basis, $\log A$ has n^2 new elements, and the last $n^2 - m$ of these are zero if and only if $\log A$ is in L. We assign these as coordinates in G_K and the conditions of the theorem are satisfied. Q.E.D.

It follows from this theorem that an analytic subgroup H of G_K has a local analytic coordinate system of dimension m and so is a p-adic Lie group.

THEOREM 7. *Let H be an analytic subgroup of G_K with Lie algebra L. Let $X^1, X^2, \cdots, X^m$ be a basis for L and $f_i(t)$ be analytic curves in H with tangents X^i respectively ($i = 1, 2, \cdots, m$). Then there exists a neighborhood of I in H all of whose elements may be put in the form*

$$f_1(t_1) \cdot f_2(t_2) \cdots f_m(t_m).$$

PROOF. Let us define the functions

$$F_i(t_1, t_2, \cdots, t_m) = [f_1(t_1) \cdot f_2(t_2) \cdots f_m(t_m)]_i, \quad i = 1, 2, \cdots, n^2,$$

i denoting the coordinate of the expression in brackets relative to a coordinate system as described in the last theorem. These give an analytic mapping of a neighborhood of the origin of K^m into a neighborhood of I in H. From the independence of the tangents to the f_i it can be shown that the Jacobian of these functions does not vanish at the origin. Hence we can solve these equations for the t_i in terms of the f_i and so by the last theorem the mapping defined by these functions covers an entire neighborhood of I in H. Q.E.D.

7. Analyticity of Subgroups

We have seen that an analytic subgroup H of G_K is closed and defined by analytic functions. It is now desirable to determine whether these two conditions are sufficient for analyticity. It will be shown that if $K = R_p$, any closed subgroup H is analytic and so is a p-adic Lie group. If $K \neq R_p$, however, this fact does not hold, as in the case of complex Lie groups. The subset of elements with coordinates in R_p is a closed subgroup but is obviously not analytic. We must therefore divide our argument into two parts.

Let H be a closed subgroup of G_K and let L be its Lie algebra. Let H' be an analytic subgroup corresponding to L and H'_r, H_r be intersections of H' and H respectively with some S_r contained in S.

LEMMA. *Let A be any element in H_r and let $X_A = \log A$. If $K = R_p$, H_r contains $\exp tX_A$ for all t in K such that $v(t) \leqq 1$. If $K \neq R_p$, but H is defined by analytic functions, the same holds.*

PROOF. i) Suppose $K = R_p$. If n is any rational integer,

$$\exp nX_A = (\exp X_A)^n$$

and so the expression on the left is in H_r, since $\exp X_A$ is in H_r. By (1), however, if t is any element of R_p such that $v(t) \leqq 1$, it is a limit of a sequence of

rational integers t_m. Hence

$$\exp tX_A = \lim_{m\to\infty} \exp t_m X_A$$

and is in H_r since H is closed.

ii) Suppose $K \neq R_p$ but that H_r is defined by analytic functions, that is, there exist analytic functions $F_i(x_1, x_2, \cdots, x_{n^2})$ such that a point A in G_K with coordinates $(a_1, a_2, \cdots, a_{n^2})$ is in H_r if and only if $F_i(a_1, a_2, \cdots, a_{n^2}) = 0$ for all i. By the above argument, $\exp tX_A$ is in H_r for all t in R_p such that $v(t) \leqq 1$. Hence we have

$$F_i[(\exp\ tX_A)_1,\ (\exp\ tX_A)_2, \cdots, (\exp\ tX_A)_{n^2}] = 0$$

for all i, and t in R_p. By (2) the F_i are analytic functions of t, and by (8) they must be zero for all t in K. Q.E.D.

THEOREM 8. *Let H satisfy the conditions in the lemma. Then H is analytic.*

PROOF. It follows from the lemma that H_r' contains H_r. By Theorem 7, a neighborhood of I in H_r' can be defined by analytic curves in H_r. This neighborhood is completely in H_r, so H is equivalent to the analytic subgroup H'. Q.E.D.

This theorem completes the setting up of the one-to-one correspondence between subgroups of G_K and subalgebras of K_{nl}.

8. Some Special Groups

Before discussing the special groups, it will be necessary to prove a theorem which occurs in the ordinary theory of Lie groups.

THEOREM 9. *Let H be an analytic subgroup of G_K with subalgebra L. Let H_1 be an invariant subgroup of H with subalgebra L_1, an ideal in L. Let $\bar{H}$ be an analytic subgroup with Lie algebra $\bar{L}$ and such that there exists a continuous homomorphism of H onto $\bar{H}$ with H_1 being the set of elements mapped on the identity I. Then $\bar{L} \cong L/L_1$.*

PROOF: We must first show that any continuous homomorphic map of a one-parameter subgroup is an analytic curve. It is sufficient to prove that any continuous homomorphic map of the ring E_K of integers of K into G_K is analytic. We need only prove this for E_{R_p} since E_K is a direct product of a finite number of the E_{R_p}. Let $t \to f(t)$ be such a mapping into some $S_r \subseteq S$ in G_K. We know that there exists an X such that

$$f(1) = \exp X,$$

and so

$$f(n) = \exp nX$$

for any rational integer n, since f is a homomorphic mapping. Since f is continuous, we have, by (1),

$$f(t) = \exp tX$$

for any t in E_{R_p}. This mapping is analytic.

It follows that if $f(t)$ is a one-parameter subgroup in H_r, its map is an analytic curve and one-parameter subgroup in $\bar{H}_r$ and all one-parameter subgroups in $\bar{H}_r$ are so obtained. This establishes a homomorphism between L and $\bar{L}$. Clearly, L_1 is the set mapped onto the zero element and so $\bar{L} = L/L_1$. Q.E.D.

Given a Lie algebra L in K_{nl}, we know that there is an infinite number of analytic subgroups of G_K whose Lie algebra is L. We are interested in finding, for a given L, an analytic subgroup H, defined by algebraic conditions, whose Lie algebra is L.

Let $\mathfrak{A}$ be an associative algebra of order n over K and $D(\mathfrak{A})$ be the Lie algebra of derivations of $\mathfrak{A}$. Let H be the group of automorphisms of $\mathfrak{A}$. H and $D(\mathfrak{A})$ may be imbedded in G_K and K_{nl} respectively.

THEOREM 10. *The Lie algebra of H is $D(\mathfrak{A})$.*

PROOF. Let X be an element of $D(\mathfrak{A})$ such that $\exp X$ is defined. Let a and b be any two elements of $\mathfrak{A}$. Then

$$\begin{aligned}[(\exp tX)\cdot a][(\exp tX)\cdot b] &= \sum_{n=0}^{\infty} t^n \left[\sum_{p=0}^{n} (1/p!(n-p)!)(X^p a)(X^{n-p} b)\right] \\ &= \sum_{n=0}^{\infty} t^n/n! \left[\sum_{p=0}^{n} C_{n,p}(X^p a)(X^{n-p} b)\right] \\ &= \sum_{n=0}^{\infty} t^n/n!\, X^n(ab) \quad {}^2 \\ &= (\exp tX)ab.\end{aligned}$$

Hence $\exp tX$ is an automorphism in $\mathfrak{A}$ for every t for which it is defined.

Conversely, let A be an element of H near I so that $A = \exp X$ for some matrix X. H is analytic, being algebraically defined, so $\exp tX$ is in H and hence is an automorphism in $\mathfrak{A}$ for all values of t for which it is defined. We have, therefore,

$$(\exp\ tX)\cdot ab = [(\exp\ tX)\cdot a][(\exp\ tX)\cdot b].$$

Expanding these series in powers of t and multiplying out,

$$ab + (tX)\cdot ab + t\epsilon_1 = ab + ta(X\cdot b) + t(X\cdot a)b + t\epsilon_2$$

and

$$X\cdot ab + \epsilon_1 = a(X\cdot b) + (X\cdot a)b + \epsilon_2 .$$

The quantities ϵ_1, ϵ_2 approach zero with t. Taking the limit of both sides as t approaches zero, we find X is a derivation. Q.E.D.

Let $\mathfrak{A}$ be an associative algebra over K and J be an involution (involutorial anti-automorphism)[3]. An element a in $\mathfrak{A}$ is called J-skew if $a^J = -a$. It is called J-orthogonal if $aa^J = a^J a = k\cdot 1$, where k is an element of K. The set $\mathfrak{S}_J$ of J-skew elements is a Lie algebra over K and the set $\mathfrak{G}_J$ of J-orthogonal elements is a multiplicative group.

[2] For this step, see Jacobson, [10], p. 207.

[3] For the required results on involutions, see Jacobson, [9], and Albert, [3], ch. X.

THEOREM 11. *The Lie algebra of* $\mathfrak{G}_J$ *is* $\mathfrak{S} = \mathfrak{S}_J \oplus K$.

PROOF. The elements of $\mathfrak{S}$ are in the form $z = a + k$, where a is in $\mathfrak{S}_J$ and k is in K. When exp z exists, it is in $\mathfrak{A}$, since $\mathfrak{A}$ has a finite basis and is closed. Hence $(\exp z)^J$ is defined. Any finite number of terms of the series $\exp z^J$ is equal to the corresponding number of terms of $(\exp z)^J$, so $\exp z^J$ exists and is equal to $(\exp z)^J$. Hence when $\exp (a + k)$ exists,

$$[\exp (a + k)][\exp (a + k)]^J = [\exp (a + k)][\exp (-a + k)]$$
$$= \exp 2k \in K,$$

and so $\exp (a + k)$ is in $\mathfrak{G}_J$.

Conversely, if $(\exp tz)(\exp tz)^J = (\exp tz)^J(\exp tz) = k \in K$, for all t in some neighborhood of zero, then $zz^J = z^Jz$ and $(\exp z)(\exp z)^J = (\exp z)(\exp z^J) = \exp (z + z^J) = k$. If z is sufficiently small, $\log k$ exists and $z + z^J = \log k = k'$. It is known that any z in $\mathfrak{A}$ can be written uniquely as

$$z = b + d \quad \text{where} \quad b^J = -b, \qquad d^J = d.$$

Hence $z + z^J = b + d + b^J + d^J = 2d = k'$. It follows that $d = k'/2$. Hence z must be in the form $a + k$ and so the Lie algebra of $\mathfrak{G}_J$ is $\mathfrak{S}$. Q.E.D.

9. Groups of Simple Lie Algebras

A simple p-adic Lie group is defined as usual as a group with no invariant subgroup which is not discrete or equivalent to the whole group. These, of course, are the groups with the simple Lie algebras.

A simple Lie algebra over any field K of characteristic zero has been shown by Landherr, [13], to be normal simple over its extended center, which is a finite algebraic extension of K. We shall, therefore, restrict ourselves to the p-adic Lie groups whose Lie algebras are the "non-exceptional" normal simple Lie algebras over K. The normal simple associative algebras over a p-adic field K have been classified by Hasse, [6]. The Lie algebras which are normal simple over any field K of characteristic zero are shown by Jacobson, [12], to arise, except for the finite number of exceptional cases, from associative algebras over K in one of the following ways:

1) (Type A_I) Let $\mathfrak{A}$ be a normal simple associative algebra over K, and let $\mathfrak{A}_l$ be the Lie algebra obtained in the usual way. The derived algebra $\mathfrak{A}_l'$ is normal simple.

2) (Types B, C, D) Let $\mathfrak{A}$ be a normal simple associative algebra over K with an involution J of first kind. Then $\mathfrak{S}_J$ is a normal simple Lie algebra.

3) (Type A_{II}) Let $\mathfrak{A}$ be a simple associative algebra over K with center $\Sigma = K(q)$, where q^2, but not q, is in K; J is an involution of second kind. The derived algebra $\mathfrak{S}_J'$ is a normal simple Lie algebra.

Using the results of Hasse, Jacobson has classified the normal simple Lie algebras over K and determined their automorphism groups. Any simple Lie algebra over K may be imbedded in K_{nl} and its automorphism group in G_K. The automorphism groups are analytic, being algebraically defined.

We wish to find for each normal simple Lie algebra L an analytic group H in G_K whose Lie algebra is L. Any other group in G_K with the same Lie algebra is equivalent to H.

We must consider separately the three cases:

1) *The group H of automorphisms of $\mathfrak{A}$ has a Lie algebra isomorphic to $\mathfrak{A}_l'$ in Case* 1.

Proof. Since $\mathfrak{A}$ is the enveloping algebra of $\mathfrak{A}_l$, and is simple, we have

$$\mathfrak{A}_l = C \oplus \mathfrak{A}_l'$$

where C is abelian and hence must be the center of $\mathfrak{A}_l$. (cf. Jacobson, [8].) C is then the center of $\mathfrak{A}$ and so is isomorphic to K. Hence

$$\mathfrak{A}_l' \cong \mathfrak{A}_l/K.$$

It is known, however, (Jacobson, [10]), that when $\mathfrak{A}$ is normal simple,

$$D(\mathfrak{A}) \cong \mathfrak{A}_l/K.$$

By Theorem 10, therefore, the Lie algebra of H is isomorphic to $\mathfrak{A}_l'$. Q.E.D.

2) *The group H of automorphisms of $\mathfrak{S}_J$ has Lie algebra $\mathfrak{S}_J$.*

Proof. Let K represent the abelian Lie algebra of scalar matrices. Let $\mathfrak{S} = \mathfrak{S}_J \oplus K$ and let $\mathfrak{G}_J$ be the group of J-orthogonal elements of $\mathfrak{A}$.

We have shown that the Lie algebra of $\mathfrak{G}_J$ is $\mathfrak{S}$. Now $\mathfrak{G}_J$ is continuously homomorphic to H with K_0 being the set of elements mapped onto I. (cf. Jacobson, [9]; K_0 is the multiplicative group of K). It is easily seen that the Lie algebra of K_0 is K, so by Theorems 9 and 11, the Lie algebra of H is isomorphic to $\mathfrak{S}/K \cong \mathfrak{S}_J$. Q.E.D.

3) *Let H be the group of automorphisms of $\mathfrak{S}_J'$ induced by inner automorphisms of $\mathfrak{A}$. The Lie algebra of H is $\mathfrak{S}_J'$.*

Proof. The enveloping algebra of $\mathfrak{S}_J$ is $\mathfrak{A}$. (cf. Jacobson, [11], p. 182). We have, therefore, as in Case 1),

$$\mathfrak{S}_J = C \oplus \mathfrak{S}_J',$$

where C is the center of the Lie algebra $\mathfrak{S}_J$. The elements of C must commute with those of $\mathfrak{S}_J$ in the associative multiplication, and hence with all of $\mathfrak{A}$, so

$$C = \mathfrak{S}_J \cap \Sigma,$$

since Σ is the center of $\mathfrak{A}$.

We have proved that the Lie algebra of $\mathfrak{G}_J$ is $\mathfrak{S} = \mathfrak{S}_J \oplus K$, and we have

$$\mathfrak{S} = \mathfrak{S}_J' \oplus (\mathfrak{S}_J \cap \Sigma) \oplus K.$$

It is known (Albert, [3]) that

$$\Sigma = K \oplus (\mathfrak{S}_J \cap \Sigma).$$

We have, therefore, $\mathfrak{S} = \mathfrak{S}_J' \oplus \Sigma$. It has been shown by Jacobson, [11], (p. 185), that $\mathfrak{G}_J$ is homomorphic to H. This can easily be seen to be continuous

and the group Σ_0 is the group mapped on the identity. It follows that H has Lie algebra $\mathfrak{S}_J'$. Q.E.D.

Although we have treated these cases separately, the result is the same in each case. Let L be the Lie algebra arising from $\mathfrak{A}$ in one of these three ways. From the results of Jacobson, [12], (pp. 339, 340), and from the fact that all automorphisms of a normal simple algebra are inner, the group H of automorphisms in L induced by inner automorphisms in $\mathfrak{A}$ has Lie algebra L in each case.

PRINCETON UNIVERSITY

BIBLIOGRAPHY

1. ADO, I. Bull. de la Soc. Physico-Math. de Kazan, **6** (1935).
2. ALBERT, A. *Modern Higher Algebra*, U. of Chicago Press, 1937.
3. ALBERT, A. *Structure of Algebras*, Am. Math. Soc. Colloquium Publications, vol. XXIV, New York, 1939.
4. CHABAUTY, C. *Sur les équations diophantiennes*, etc., Annali di Matematica, **17** (1938), pp. 127–168.
5. CHEVALLEY, C. *Sur la théorie du corps de classes*, etc., Journal of the Faculty of Science, Tokyo, **2** (1933), pp. 365–474.
6. HASSE, H. *Über p-adische Schiefkörper*, Math. Annalen, **104** (1931), pp. 495–534.
7. HAUSDORFF, F. *Die symbolische Exponentialformel*, etc., Ber. der Ges. der Wiss. zu Leipzig, **58** (1906), pp. 19–48.
8. JACOBSON, N. *Rational Methods in the Theory of Lie Algebras*, these Annals, **36** (1935), pp. 875–881.
9. JACOBSON, N. *A Class of Normal Simple Lie Algebras of Characteristic Zero*, these Annals, **38** (1937), pp. 508–517.
10. JACOBSON, N. *Abstract Derivation and Lie Algebras*, Trans. Am. Math. Soc., **42** (1937), pp. 206–224.
11. JACOBSON, N. *Simple Lie Algebras of Type* A, these Annals, **39** (1938), pp. 181–188.
12. JACOBSON, N. *Simple Lie Algebras over a Field of Characteristic Zero*, Duke Math. J., **4** (1938), pp. 534–551.
13. LANDHERR, W. *Über einfache Liesche Ringe*, Hamb. Abhandlungen, **11** (1935), pp. 41–64.
14. LUTZ, E. *Sur l'equation* $y^2 = x^3 - Ax - B$, etc., J. für die reine und angew. Math., **177** (1937), pp. 238–247.
15. PONTRJAGIN, L. *Topological Groups*, Princeton, 1939.

Annals of Mathematics
Vol. 43, No. 4, October, 1942

ON THE MODULAR REPRESENTATIONS OF THE SYMMETRIC GROUP

By R. M. Thrall and C. J. Nesbitt

(Received December 11, 1941)

1. Introduction

The purpose of the present paper is to determine all modular (matrix) representations of the symmetric group, $\mathfrak{S}_m$, of degree m, where $m < 2p$. The elements of the representing matrices are to be chosen from any field, $\mathfrak{k}$, of characteristic p. Every indecomposable (modular) representation of $\mathfrak{S}_m$ is equivalent to a rational one (i.e. to one in which $\mathfrak{k}$ is the prime field) so the nature of the field, $\mathfrak{k}$, is of no particular interest in what follows.

In the last several years considerable progress has been made in the theory of modular representations of finite groups.[1] This theory is especially well worked out in case the order of the group is divisible by only the first power of p; hence our requirement $m < 2p$. (Actually we treat here only the cases $p \leqq m < 2p$ since for $m < p$ the ordinary theory applies, leaving no problem.)

Any representation of a group (or of an algebra) is completely characterized by its indecomposable constituents. In general there are an infinite number of inequivalent indecomposable modular representations of a finite group. One of the main results of the present paper is a proof that the symmetric group of degree less than $2p$ has only a finite number of inequivalent indecomposable representations. In sections 2–4 we determine the structure of the regular representation of $\mathfrak{S}_m$ (or of its $\mathfrak{k}$ group ring $\mathfrak{R}_m$). In section 5 we show how any indecomposable representation can be built up from "elementary modules" (see section 3 for definition and references) of the group ring.

In section 6 we state Nakayama's results[2] on the modular representations of $\mathfrak{S}_m$ and add a discussion of the behavior of representations of $\mathfrak{S}_m$ when considered only for elements of $\mathfrak{S}_{m-1}$.

The final three sections are devoted to specific determination of all representations of $\mathfrak{S}_p$ with indications of generalization to $\mathfrak{S}_{p+1}$, $\mathfrak{S}_{p+2}$.

2. Preliminaries

Let $m = p + l, l < p$, and let

$$(1) \qquad m = \alpha_1 + 2\alpha_2 + \cdots + m\alpha_m$$

be a partition of m. Corresponding to this partition (α) there is a class $C(\alpha)$ of conjugate elements of $\mathfrak{S}_m$ such that if $s \in C(\alpha)$, then s is a product of α_1 1-cycles,

---

[1] See the bibliography at the end of the paper.
[2] See [8] especially part II.

α_2 2-cycles, $\cdots$, α_m m-cycles. The number of conjugate classes, and hence the number of ordinary irreducible representations is P_m, the number of partitions of m.

Since $m = p + l$, α_p in (1) may be 0 or 1. The class $C(\alpha)$ is p-singular (that is, the order of the elements of $C(\alpha)$ is divisible by p) if and only if $\alpha_p = 1$. Then the number of p-singular classes is equal to P_l. It follows that the *number of modular irreducible representations which is equal to the number of p-regular classes is* $P_m - P_l$.[3]

Corresponding to the decomposition of the group ring $\mathfrak{R}_m$ into a direct sum of directly indecomposable invariant subalgebras there exists a classification of the ordinary irreducible and the modular irreducible representations, and their characters into blocks.[4] A block $\mathfrak{B}_\tau$ is said to be of type β if all the ordinary irreducible representations which belong to $\mathfrak{B}_\tau$ have degrees $\equiv 0 \pmod{p^\beta}$, but at least one of these degrees $\not\equiv 0 \pmod{p^{\beta+1}}$. In our case we have just blocks of type 0 or lowest kind, and blocks of highest kind (here of type 1).

Theorem II of [1] states that the number t_0 of blocks of lowest kind is equal to the number of p-regular classes of conjugate elements where the number of elements in the class is prime to p. The number of elements in the class $C(\alpha)$ is $m!/n(\alpha)$ where $n(\alpha) = \alpha_1!\alpha_2!2^{\alpha_2} \cdots \alpha_m!m^{\alpha_m}$ is the order of the normalizor of any element s in $C(\alpha)$. To determine t_0 for our case, we count the partitions of m where $\alpha_p = 0$ (that is, $C(\alpha)$ is p-regular) and $m!/n(\alpha) \not\equiv 0 \pmod p$. Since for $i > 1$, α_i is less than p, and α_p is here 0, then $n(\alpha) \equiv 0 \pmod p$ only if $\alpha_1 \geqq p$. One easily sees a 1–1 correspondence between the partitions of m with $\alpha_1 \geqq p$, and the partitions of l. We thus obtain $t_0 = P_l$.

We denote by x_τ, y_τ the numbers of ordinary and of modular irreducible characters which belong to a block $\mathfrak{B}_\tau$. If $\mathfrak{B}_\tau$ is of highest kind $x_\tau = y_\tau = 1$; for $\mathfrak{B}_\tau$ of lowest kind $x_\tau \geqq y_\tau + 1$.[5] But, by the above, $\sum x_\tau - \sum y_\tau = P_m - (P_m - P_l) = P_l$, and there are P_l blocks of lowest kind, so the only possibility is that for each block of lowest kind $x_\tau = y_\tau + 1$. We shall show below that $x_\tau = p$ for blocks of lowest kind.

In the theory of modular representations of groups two sets of numbers play leading roles: the decomposition numbers describe the splitting of the ordinary irreducible representations (when taken in the modular sense) into modular irreducible constituents; the Cartan invariants give the multiplicities of the modular irreducible representations as constituents of the indecomposable parts of the modular regular representation. If D_τ, C_τ denote the matrices of decomposition and Cartan numbers for the block B_τ, then a main theorem is that[6]

$$C_\tau = D'_\tau D_\tau. \tag{2}$$

[3] Cf. Brauer-Nesbitt [1], §8 for proof that the number of p-regular classes is equal the number of modular irreducible representations.

[4] Brauer-Nesbitt [1], §9, and Nakayama [9], Theorem 5.

[5] Brauer-Nesbitt [1], §19, Theorem 5.

[6] Brauer-Nesbitt [1], §§4, 5 and 9.

From Brauer's work (in particular, see Theorem 14 of [2]) we have in our case that for a block of lowest kind

$$(3) \qquad D_\tau = \left\| \begin{matrix} 1 & & & \\ 1 & 1 & & \\ & 1 & 1 & \\ & & \cdots\cdots & \\ & & 1 & 1 \\ & & & 1 \end{matrix} \right\| .$$

Here D_τ has x_τ rows, y_τ columns. Then C_τ has y_τ rows, y_τ columns, and from (2), (3)

$$(4) \qquad C_\tau = \left\| \begin{matrix} 2 & 1 & & & \\ 1 & 2 & 1 & & \\ & 1 & 2 & 1 & \\ & & \cdots\cdots & & \\ & & 1 & 2 & 1 \\ & & & 1 & 2 \end{matrix} \right\| .$$

It follows from (4) that det C_τ is $y_\tau + 1 = x_\tau$.

We consider now the set M of $n(\alpha)$'s corresponding to the p-regular classes $C(\alpha)$, and form the product $\prod n(\alpha)$. For each of the P_l partitions (α) with $\alpha_p = 0, \alpha_1 \geqq p, n(\alpha)$ is divisible by p, and all other $n(\alpha)$ of the set $M \not\equiv 0 \pmod p$. Then p^{P_l} is the highest power of p which divides $\prod n(\alpha)$. By Theorem I of [3] the determinant of the complete matrix C of Cartan invariants is then equal to p^{P_l}, that is,

$$\det C = \prod \det C_\tau = p^{P_l}.$$

But for blocks B_τ of highest kind det $C_\tau = 1$, and for blocks of lowest kind det $C_\tau = x_\tau$, and there are P_l blocks of lowest kind; hence for each such block $x_\tau = p$. *To each block of lowest kind there belong p ordinary irreducible representations and $p - 1$ modular irreducible representations.*

It follows also that there are $P_m - pP_l$ blocks of highest kind and that the total number of blocks is $P_m - (p - 1)P_l$.

3. Loewy form

Let $\mathfrak{A}$ denote any matrix representation of $\mathfrak{R}_m$. We consider the algebra $\mathfrak{A}$ as a system of linear transformations of a vector space $\mathfrak{B}$ of suitable dimension. Let $\mathfrak{N}, \mathfrak{N}^2, \cdots, \mathfrak{N}^{t-1}, \mathfrak{N}^t = (0)$ denote the powers of the radical of $\mathfrak{A}$. We form the upper Loewy series of $\mathfrak{B}$, namely[7]

$$(5) \qquad \mathfrak{B} \supset \mathfrak{N}\mathfrak{B} \supset \mathfrak{N}^2\mathfrak{B} \cdots \supset \mathfrak{N}^{t-1}\mathfrak{B} \supset 0.$$

[7] For a discussion of Loewy series see §§2, 5 of [4]. That (5) is the upper series for $\mathfrak{B}$ follows by proving Theorem 12.3A of [4] for vector spaces having $\mathfrak{A}$ as operator system rather than for ideals of $\mathfrak{A}$.

Here $\mathfrak{N}^{\rho-1}\mathfrak{B}/\mathfrak{N}^{\rho}\mathfrak{B}$ is the unique maximal completely reducible factor group that may be obtained from $\mathfrak{N}^{\rho-1}\mathfrak{B}(\rho = 1, 2, \cdots, t)$. If we adapt the coordinate system in $\mathfrak{B}$ to the series (5), then $\mathfrak{A}$ appears in upper Loewy form,

$$\mathfrak{A} \sim \left\| \begin{matrix} \mathfrak{L}_1(\mathfrak{A}) & & \\ & \mathfrak{L}_2(\mathfrak{A}) & \\ * & & \\ & & \mathfrak{L}_t(\mathfrak{A}) \end{matrix} \right\| \tag{6}$$

where the $\mathfrak{L}_i(\mathfrak{A})$ are the upper Loewy constituents of $\mathfrak{A}$, and are completely reducible. t may be called the Loewy length of $\mathfrak{A}$.[8]

We consider now the Loewy form of the indecomposable parts of the regular representation $\mathfrak{R}$ of $\mathfrak{R}_m$. Let $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_k$ denote the modular irreducible representations of $\mathfrak{R}_m$. It follows from [6] (see, in particular, Theorems 2, 3, and 8) that we may denote the indecomposable parts of $\mathfrak{R}$ by $\mathfrak{U}_1, \cdots, \mathfrak{U}_k$ where, if $\mathfrak{U}_\kappa$ is written in the form (6), $\mathfrak{L}_1(\mathfrak{U}_\kappa) = \mathfrak{L}_t(\mathfrak{U}_\kappa) = \mathfrak{F}_\kappa$

$$\mathfrak{U}_\kappa = \left\| \begin{matrix} \mathfrak{F}_\kappa & & & \\ & \mathfrak{L}_2(\mathfrak{U}_\kappa) & & \\ & & \ddots & \\ * & & \mathfrak{L}_{t-1}(\mathfrak{U}_\kappa) & \\ & & & \mathfrak{F}_\kappa \end{matrix} \right\|. \tag{7}$$

We have considered here only the upper Loewy forms. We might similarly have discussed the lower Loewy forms. For the indecomposable parts $\mathfrak{U}_\kappa$ of $\mathfrak{R}_m$ (see below) the upper Loewy forms coincide with the lower Loewy forms.

In the following we shall use the notation $c_{\kappa\lambda}$ to denote the multiplicity of $\mathfrak{F}_\lambda$ as constituent of $\mathfrak{U}_\kappa$; $c_{\kappa\lambda}$ is a Cartan invariant.

4. Elementary modules

Let $\mathfrak{R}$ denote the modular regular representation of the group ring $\mathfrak{R}_m$ of $\mathfrak{S}_m$. It is well known that $\mathfrak{R}$ is a faithful representation of $\mathfrak{R}_m$. Then instead of considering elements of $\mathfrak{R}_m$ we shall for the time being speak of the corresponding elements of $\mathfrak{R}$. We assume $\mathfrak{R}$ to be in reduced form, that is

$$\mathfrak{R} = \left\| \begin{matrix} \mathfrak{R}_{11} & & \\ \mathfrak{R}_{21} & \mathfrak{R}_{22} & \\ \vdots & & \ddots \\ \mathfrak{R}_{s1} & \cdots & \mathfrak{R}_{ss} \end{matrix} \right\|$$

where the $\mathfrak{R}_{ii}$ denote irreducible constituents of $\mathfrak{R}$. We may further assume that $\mathfrak{R} = \mathfrak{R}^* + \mathfrak{N}$, where $\mathfrak{R}^*$ is the semi-simple algebra obtained from $\mathfrak{R}$ by replacing the $\mathfrak{R}_{ij}$ with $i > j$ by 0, and where $\mathfrak{N}$, the radical of $\mathfrak{R}$, has 0 in place

---

[8] Cf. Brauer, [4], §5.

From Brauer's work (in particular, see Theorem 14 of [2]) we have in our case that for a block of lowest kind

$$(3) \qquad D_\tau = \left\| \begin{array}{l} 1 \\ 11 \\ \;\;11 \\ \cdots\cdots\cdots \\ \qquad\;\; 11 \\ \qquad\quad\;\; 1 \end{array} \right\| .$$

Here D_τ has x_τ rows, y_τ columns. Then C_τ has y_τ rows, y_τ columns, and from (2), (3)

$$(4) \qquad C_\tau = \left\| \begin{array}{l} 21 \\ 121 \\ \;\;121 \\ \cdots\cdots \\ \qquad\;\; 121 \\ \qquad\quad\;\; 12 \end{array} \right\| .$$

It follows from (4) that det C_τ is $y_\tau + 1 = x_\tau$.

We consider now the set M of $n(\alpha)$'s corresponding to the p-regular classes $C(\alpha)$, and form the product $\prod n(\alpha)$. For each of the P_l partitions (α) with $\alpha_p = 0, \alpha_1 \geqq p, n(\alpha)$ is divisible by p, and all other $n(\alpha)$ of the set $M \not\equiv 0 \pmod{p}$. Then p^{P_l} is the highest power of p which divides $\prod n(\alpha)$. By Theorem I of [3] the determinant of the complete matrix C of Cartan invariants is then equal to p^{P_l}, that is,

$$\det C = \prod \det C_\tau = p^{P_l} .$$

But for blocks B_τ of highest kind det $C_\tau = 1$, and for blocks of lowest kind det $C_\tau = x_\tau$, and there are P_l blocks of lowest kind; hence for each such block $x_\tau = p$. *To each block of lowest kind there belong p ordinary irreducible representations and p − 1 modular irreducible representations.*

It follows also that there are $P_m - pP_l$ blocks of highest kind and that the total number of blocks is $P_m - (p-1)P_l$.

3. Loewy form

Let $\mathfrak{A}$ denote any matrix representation of $\mathfrak{R}_m$. We consider the algebra $\mathfrak{A}$ as a system of linear transformations of a vector space $\mathfrak{B}$ of suitable dimension. Let $\mathfrak{N}, \mathfrak{N}^2, \cdots, \mathfrak{N}^{t-1}, \mathfrak{N}^t = (0)$ denote the powers of the radical of $\mathfrak{A}$. We form the upper Loewy series of $\mathfrak{B}$, namely[7]

$$(5) \qquad \mathfrak{B} \supset \mathfrak{N}\mathfrak{B} \supset \mathfrak{N}^2\mathfrak{B} \cdots \supset \mathfrak{N}^{t-1}\mathfrak{B} \supset 0.$$

[7] For a discussion of Loewy series see §§2, 5 of [4]. That (5) is the upper series for $\mathfrak{B}$ follows by proving Theorem 12.3A of [4] for vector spaces having $\mathfrak{A}$ as operator system rather than for ideals of $\mathfrak{A}$.

Here $\mathfrak{N}^{\rho-1}\mathfrak{B}/\mathfrak{N}^{\rho}\mathfrak{B}$ is the unique maximal completely reducible factor group that may be obtained from $\mathfrak{N}^{\rho-1}\mathfrak{B}(\rho = 1, 2, \cdots, t)$. If we adapt the coordinate system in $\mathfrak{B}$ to the series (5), then $\mathfrak{A}$ appears in upper Loewy form,

$$(6) \qquad \mathfrak{A} \sim \begin{Vmatrix} \mathfrak{L}_1(\mathfrak{A}) & & \\ & \mathfrak{L}_2(\mathfrak{A}) & \\ * & & \ddots \\ & & & \mathfrak{L}_t(\mathfrak{A}) \end{Vmatrix}$$

where the $\mathfrak{L}_i(\mathfrak{A})$ are the upper Loewy constituents of $\mathfrak{A}$, and are completely reducible. t may be called the Loewy length of $\mathfrak{A}$.[8]

We consider now the Loewy form of the indecomposable parts of the regular representation $\mathfrak{R}$ of $\mathfrak{R}_m$. Let $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_k$ denote the modular irreducible representations of $\mathfrak{R}_m$. It follows from [6] (see, in particular, Theorems 2, 3, and 8) that we may denote the indecomposable parts of $\mathfrak{R}$ by $\mathfrak{U}_1, \cdots, \mathfrak{U}_k$ where, if $\mathfrak{U}_\kappa$ is written in the form (6), $\mathfrak{L}_1(\mathfrak{U}_\kappa) = \mathfrak{L}_t(\mathfrak{U}_\kappa) = \mathfrak{F}_\kappa$

$$(7) \qquad \mathfrak{U}_\kappa = \begin{Vmatrix} \mathfrak{F}_\kappa & & & \\ & \mathfrak{L}_2(\mathfrak{U}_\kappa) & & \\ & & \ddots & \\ * & & & \mathfrak{L}_{t-1}(\mathfrak{U}_\kappa) & \\ & & & & \mathfrak{F}_\kappa \end{Vmatrix}.$$

We have considered here only the upper Loewy forms. We might similarly have discussed the lower Loewy forms. For the indecomposable parts $\mathfrak{U}_\kappa$ of $\mathfrak{R}_m$ (see below) the upper Loewy forms coincide with the lower Loewy forms.

In the following we shall use the notation $c_{\kappa\lambda}$ to denote the multiplicity of $\mathfrak{F}_\lambda$ as constituent of $\mathfrak{U}_\kappa$; $c_{\kappa\lambda}$ is a Cartan invariant.

4. Elementary modules

Let $\mathfrak{R}$ denote the modular regular representation of the group ring $\mathfrak{R}_m$ of $\mathfrak{S}_m$. It is well known that $\mathfrak{R}$ is a faithful representation of $\mathfrak{R}_m$. Then instead of considering elements of $\mathfrak{R}_m$ we shall for the time being speak of the corresponding elements of $\mathfrak{R}$. We assume $\mathfrak{R}$ to be in reduced form, that is

$$\mathfrak{R} = \begin{Vmatrix} \mathfrak{R}_{11} & & \\ \mathfrak{R}_{21} & \mathfrak{R}_{22} & \\ \vdots & & \ddots \\ \mathfrak{R}_{s1} & \cdots & \mathfrak{R}_{ss} \end{Vmatrix}$$

where the $\mathfrak{R}_{ii}$ denote irreducible constituents of $\mathfrak{R}$. We may further assume that $\mathfrak{R} = \mathfrak{R}^* + \mathfrak{N}$, where $\mathfrak{R}^*$ is the semi-simple algebra obtained from $\mathfrak{R}$ by replacing the $\mathfrak{R}_{ij}$ with $i > j$ by 0, and where $\mathfrak{N}$, the radical of $\mathfrak{R}$, has 0 in place

[8] Cf. Brauer, [4], §5.

of the $\mathfrak{R}_{ii}$ in the main diagonal. Let, as before, $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_k$ denote the distinct modular irreducible representations of $\mathfrak{R}_m$, and $f_1, f_2, \cdots, f_k$ their degrees. We mean by $\mathfrak{R}_\kappa^*$ the simple subalgebra of $\mathfrak{R}^*$ which has 0 in place of all $\mathfrak{R}_{ii}$ except those which are equivalent to $\mathfrak{F}_\kappa$. Let $e_\kappa(ij)$ $(i, j = 1, 2, \cdots, f_\kappa)$ be a set of matrix units for $\mathfrak{R}_\kappa^*$. We denote the unit element of the simple algebra $\mathfrak{R}_\kappa^*$ by $e_\kappa = \sum_{i=1}^{f_\kappa} e_\kappa(ii)$. An element a of $\mathfrak{R}$ we say is of type (κ, λ) if $e_\kappa a e_\lambda = a$. In [6] it was shown that a system of *primitive* elements

$$(8) \qquad b_1, b_2, \cdots, b_m, \qquad m = \sum_{\kappa,\lambda=1}^{k} c_{\kappa\lambda}$$

could be chosen so that these and their right and left products with suitable $e_\kappa(ij)$ gave a basis for $\mathfrak{R}$. More fully, for each type (κ, λ) there are $c_{\kappa\lambda}$ elements b in the set (8) of that type. If b_ρ is of type (κ, λ) then we take all the elements

$$(9) \qquad e_\kappa(i1)b_\rho e_\lambda(1j) \qquad (i = 1, 2, \cdots, f_\kappa, \ j = 1, 2, \cdots, f_\lambda)$$

for part of our basis of $\mathfrak{R}$. Doing this for each b_ρ we obtain what has been called the Cartan basis of $\mathfrak{R}$. This Cartan basis can be so arranged that the regular representation with respect to this new basis is split into indecomposable and irreducible parts.

The Cartan basis system is the starting point for the definition recently given by W. M. Scott of *elementary modules*.[9] An element a of $\mathfrak{R}$, expressed in terms of the Cartan basis elements will have the form

$$a = \sum_{\rho,i,j} h_{ij}^\rho(a) e_\kappa(i1) b_\rho e_\lambda(1j).$$

Scott arranges the coefficients $h_{ij}^\rho(a)$, for a fixed ρ, in a matrix $H_\rho(a) = \| h_{ij}^\rho(a) \|$. The Abelian additive group generated by the matrices $H^\rho(a)$, $a \in \mathfrak{R}$, Scott has called an elementary module. The significance of these for us here is that the representations of $\mathfrak{R}_m$, in particular, the regular representations, may be expressed in simple form by these elementary modules.

We have two cases to consider:

(1) *Blocks of highest kind.* Let $\mathfrak{B}_\tau$ be a block of highest kind and $\mathfrak{F}_\lambda$ be the unique modular irreducible representation belonging to $\mathfrak{B}_\tau$. Then $c_{\lambda\mu} = 0$ for $\lambda \neq \mu$, $c_{\lambda\lambda} = 1$, and the elements $e_\lambda(ij)$ $(i, j = 1, 2, \cdots, f_\lambda)$ may be taken as the Cartan basis for $\mathfrak{B}_\tau$. The corresponding elementary module is just $\mathfrak{F}_\lambda$. Further, $\mathfrak{F}_\lambda = \mathfrak{U}_\lambda$, where $\mathfrak{U}_\lambda$ is the unique indecomposable part of $\mathfrak{R}$ corresponding to $\mathfrak{F}_\lambda$.

(2) *Blocks of lowest kind.* Let $\mathfrak{B}_\tau$ be a block of lowest kind, and let us choose our enumeration of the modular irreducible representations so that $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_{p-1}$ belong to $\mathfrak{B}_\tau$, and further so that the matrix C_τ of Cartan invariants for $\mathfrak{B}_\tau$ is in the form (4).

Then for $\kappa \neq 1$ or $p - 1$, $c_{\kappa\kappa} = 2$, $c_{\kappa-1,\kappa} = 1 = c_{\kappa+1,\kappa}$, and all other $c_{\lambda\kappa}$ are zero. Corresponding to these invariants there exist primitive elements $e_\kappa(11)$ and $b_{\kappa\kappa}$ of type (κ, κ), $b_{\kappa-1,\kappa}$ of type $(\kappa - 1, \kappa)$ and $b_{\kappa+1,\kappa}$ of type $(\kappa + 1, \kappa)$.

[9] Scott, [10].

The Cartan basis elements

$$
\begin{array}{llll}
 & \text{(a)} & e_\kappa(j1) & j = 1, 2, \cdots, f_\kappa \\
 & & e_{\kappa-1}(k1)b_{\kappa-1,\kappa} & k = 1, 2, \cdots, f_{\kappa-1} \\
(10) & \text{(b)} & & \\
 & & e_{\kappa+1}(l1)b_{\kappa+1,\kappa} & l = 1, 2, \cdots, f_{\kappa+1} \\
 & \text{(c)} & e_\kappa(m1)b_{\kappa\kappa} & m = 1, 2, \cdots, f_\kappa
\end{array}
$$

form a basis for an indecomposable $\mathfrak{R}$-left ideal $\mathfrak{l}_\kappa = \mathfrak{R}e_\kappa(11)$ which is a summand in the expression of $\mathfrak{R}$ as a direct sum of indecomposable left ideals. $\mathfrak{l}_\kappa$ may be considered as the representation space of the indecomposable part $\mathfrak{U}_\kappa$ of $\mathfrak{R}$. The form (7) of $\mathfrak{U}_\kappa$ shows that the Loewy length t of $\mathfrak{U}_\kappa \geqq 3$. It cannot be greater than 3 for then $\mathfrak{N}^3\mathfrak{l}_\kappa \neq 0$, that is, there would have to exist primitive elements of type $(*, \kappa)$ belonging to $\mathfrak{N}^3$. Then t must equal 3, and again from (7) it follows that $\mathfrak{N}^2\mathfrak{l}_\kappa$ is generated by the elements (c), $b_{\kappa\kappa}$ belongs to $\mathfrak{N}^2$, and that $b_{\kappa-1,\kappa}$, $b_{\kappa+1,\kappa}$ belong to $\mathfrak{N}$. We may assume the primitive elements so chosen that $b_{\kappa,\kappa-1}b_{\kappa-1,\kappa} = b_{\kappa,\kappa+1}b_{\kappa+1,\kappa} = b_{\kappa,\kappa}$ (here $b_{\kappa,\kappa-1}$, $b_{\kappa,\kappa+1}$ correspond to the Cartan invariants $c_{\kappa,\kappa-1} = c_{\kappa,\kappa+1} = 1$). We denote by $\mathfrak{F}_\kappa$, $\mathfrak{H}_\kappa^{\kappa-1}$, $\mathfrak{H}_\kappa^{\kappa+1}$, $\mathfrak{H}_\kappa^\kappa$ the elementary modules corresponding to $e_\kappa(11)$, $b_{\kappa-1,\kappa}$, $b_{\kappa+1,\kappa}$ and $b_{\kappa,\kappa}$. Then following the method of [6], §3, taking in our basis of $\mathfrak{l}_\kappa$ first the elements (a), then those of (b) and lastly those of (c) we calculate $\mathfrak{U}_\kappa$ to be

$$
\mathfrak{U}_\kappa = \left\| \begin{array}{llll}
\mathfrak{F}_\kappa & & & \\
\mathfrak{H}_\kappa^{\kappa-1} & \mathfrak{F}_{\kappa-1} & & \\
\mathfrak{H}_\kappa^{\kappa+1} & 0 & \mathfrak{F}_{\kappa+1} & \\
\mathfrak{H}_\kappa^\kappa & \mathfrak{H}_{\kappa-1}^\kappa & \mathfrak{H}_{\kappa+1}^\kappa & \mathfrak{F}_\kappa
\end{array} \right\|. \tag{11}
$$

We similarly can compute

$$
\mathfrak{U}_1 = \left\| \begin{array}{lll}
\mathfrak{F}_1 & & \\
\mathfrak{H}_1^2 & \mathfrak{F}_2 & \\
\mathfrak{H}_1^1 & \mathfrak{H}_2^1 & \mathfrak{F}_1
\end{array} \right\|
$$

$$
\mathfrak{U}_{p-1} = \left\| \begin{array}{lll}
\mathfrak{F}_{p-2} & & \\
\mathfrak{H}_{p-1}^{p-2} & \mathfrak{F}_{p-2} & \\
\mathfrak{H}_{p-1}^{p-1} & \mathfrak{H}_{p-2}^{p-1} & \mathfrak{F}_{p-1}
\end{array} \right\|.
$$

5. Indecomposable representations of $\mathfrak{R}_m$

We set out now to determine all indecomposable representations of $\mathfrak{R}_m$. A first clue is that the Loewy length $l(\mathfrak{M})$ of any representation $\mathfrak{M}$ of $\mathfrak{R}_m \leqq 3$. This follows from our above result that $l(\mathfrak{U}_\kappa) = 1$ or 3 according as $\mathfrak{U}_\kappa$ belongs to a block of highest or of lowest kind, and Theorems 6.6C, 11.5B of [3].

A second simplification comes from observing that $\mathfrak{M}$ may be taken in upper

Loewy form and at the same time have its simple parts expressed in terms of the elementary modules $\mathfrak{H}$. Let us picture $\mathfrak{M}$ in reduced form

$$\mathfrak{M} = \left\Vert \begin{array}{ll} \mathfrak{M}_{11} & \\ \mathfrak{M}_{21} & \mathfrak{M}_{22} \\ \cdots & \cdots \end{array} \right\Vert$$

and denote by $\mathfrak{M}^*$ the subalgebra of $\mathfrak{M}$ obtained by replacing in $\mathfrak{M}$ the parts $\mathfrak{M}_{ij}$, $j < i$ by 0. It follows from Scott's results,[10] that if in the representation $\mathfrak{M}$, the semisimple algebra $\mathfrak{R}^*$ of $\mathfrak{R}$ is mapped into $\mathfrak{M}^*$, then the simple parts $\mathfrak{M}_{ij}$ are expressible as linear combinations of the elementary modules $\mathfrak{H}$. Let $\mathfrak{V}$ be the representation space of $\mathfrak{M}$, and suppose the Loewy length of $\mathfrak{M}$ is 3. We take the Loewy series $\mathfrak{V} \supset \mathfrak{N}\mathfrak{V} \supset \mathfrak{N}^2\mathfrak{V} \supset 0$. Here $\mathfrak{N}^2\mathfrak{V}$ may be considered as an $\mathfrak{R}^*$ space, and as such is a direct sum of irreducible $\mathfrak{R}^*$ spaces, $\mathfrak{N}^2\mathfrak{V} = \mathfrak{V}_1^{(2)} \oplus \cdots \oplus \mathfrak{V}_{\alpha_2}^{(2)}$. Further $\mathfrak{N}\mathfrak{V}$ as an $\mathfrak{R}^*$ space is a direct sum of $\mathfrak{N}^2\mathfrak{V}$ and a complementary space which may also be written as a direct sum of irreducible $\mathfrak{R}^*$ spaces. By continuing the argument, we obtain that as an $\mathfrak{R}^*$ space

$$\mathfrak{V} = \mathfrak{V}_1^{(0)} \oplus \cdots \oplus \mathfrak{V}_{\alpha_0}^{(0)} \oplus \mathfrak{V}_1^{(1)} \oplus \cdots \oplus \mathfrak{V}_{\alpha_1}^{(1)} \oplus \mathfrak{V}_1^{(2)} \oplus \cdots \oplus \mathfrak{V}_{\alpha_2}^{(2)}.$$

Adapting the co-ordinate system to this decomposition of $\mathfrak{V}$, we obtain $\mathfrak{R}^*$ mapped on $\mathfrak{M}^*$ in $\mathfrak{M}$, and simultaneously have $\mathfrak{M}$ in Loewy form.

Let now $\mathfrak{M}$ be a modular indecomposable representation of $\mathfrak{R}_m$ and let $\mathfrak{R}_m = \mathfrak{T}_1 \oplus \mathfrak{T}_2 \oplus \cdots \oplus \mathfrak{T}_q$ denote the decomposition of $\mathfrak{R}_m$ into indecomposable two-sided ideals. Since the representation space may be written

$$\mathfrak{V} = \mathfrak{R}_m\mathfrak{V} = \mathfrak{T}_1\mathfrak{V} \oplus \mathfrak{T}_2\mathfrak{V} \oplus \cdots \oplus \mathfrak{T}_q\mathfrak{V}$$

and $\mathfrak{M}$ is indecomposable, we find that only one $\mathfrak{T}_\sigma\mathfrak{V}$, say $\mathfrak{T}_i\mathfrak{V}$, can be different from zero. This implies that only $\mathfrak{T}_i$ is mapped on something different from zero in the representation $\mathfrak{M}$, and that $\mathfrak{M}$ contains only modular irreducible representations belonging to the block corresponding to $\mathfrak{T}_i$.

We use the Loewy length $l(\mathfrak{M})$ of $\mathfrak{M}$ to distinguish three cases.

1) $l(\mathfrak{M}) = 1$. Then $\mathfrak{M}$ is both completely reducible and indecomposable, and so $\mathfrak{M}$ must be equivalent to a modular irreducible representation $\mathfrak{F}_\kappa$. We observe that if $\mathfrak{M}$ contains a modular irreducible constituent belonging to a block of highest kind then $l(\mathfrak{M}) = 1$.[11]

2) $l(\mathfrak{M}) = 2$. Then the modular irreducible constituents of $\mathfrak{M}$ must belong to a block $\mathfrak{B}_\tau$ of lowest kind. Let $\mathfrak{F}_1$, $\mathfrak{H}_1^1$, $\mathfrak{H}_2^1$, $\mathfrak{H}_1^2$, $\mathfrak{F}_2$, $\mathfrak{H}_2^2$, $\mathfrak{H}_3^2$, $\cdots$, $\mathfrak{H}_{p-2}^{p-1}$, $\mathfrak{F}_{p-1}$, $\mathfrak{H}_{p-1}^{p-1}$ denote the elementary modules of $\mathfrak{B}_\tau$; here $\mathfrak{F}_1, \cdots, \mathfrak{F}_{p-1}$ are the modular irreducible representations of $\mathfrak{B}_\tau$, $\mathfrak{H}_\lambda^\kappa$, $\lambda \neq \kappa$, are the elementary modules of $\mathfrak{B}_\rho$ which belong to the first power $\mathfrak{N}$ of the radical, and the $\mathfrak{H}_\kappa^\kappa$ are those which belong to $\mathfrak{N}^2$. As here the Loewy series for $\mathfrak{M}$ is $\mathfrak{V} \supset \mathfrak{N}\mathfrak{V} \supset (0)$,

---

[10] Cf. Scott, [10].

[11] For a modular irreducible representation belonging to a block of highest kind is also an indecomposable part of the regular representation. Then apply Remark 3 of [7].

the $\mathfrak{H}_\kappa^\kappa$ will not appear in $\mathfrak{M}$. Our notation for the elementary modules is based on the form (4) of the matrix C_r of Cartan invariants. From the above considerations, we may suppose that $\mathfrak{M}$ in Loewy form is written (denoting the unit matrix of degree f by E_f)

$$(12)\quad \mathfrak{M} = \left\| \begin{array}{cccccc} E_{h_1} \times \mathfrak{F}_1 & & & & & \\ & E_{h_3} \times \mathfrak{F}_3 & & & & \\ & & \ddots & & & \\ & & E_{h_{p-2}} \times \mathfrak{F}_{p-2} & & & \\ A_1^2 \times \mathfrak{H}_1^2, & A_3^2 \times \mathfrak{H}_3^2 & & E_{h_2} \times \mathfrak{F}_2 & & \\ & A_3^4 \times \mathfrak{H}_3^4, . & & & E_{h_4} \times \mathfrak{F}_4 & \\ & \ddots & & & \ddots & \\ & & A_{p-2}^{p-1} \times \mathfrak{H}_{p-2}^{p-1} & & & E_{h_{p-1}} \times \mathfrak{F}_{p-1} \end{array} \right\|$$

or in a similar form with the even $\mathfrak{F}_\kappa$ in the top Loewy constituent and the odd $\mathfrak{F}_\kappa$ in the bottom constituent. If we assumed that both even and odd constituents $\mathfrak{F}_\kappa$ appear in the same Loewy constituent, then by permutation of the rows and columns in $\mathfrak{M}$ a decomposition of $\mathfrak{M}$ would be obtained.

From Schur's lemma it follows that a matrix P which commutes with $\mathfrak{M}$ has the form

$$(13)\qquad P = \left\| \begin{array}{ccc} P_1 \times E_{h_1} & & \\ & P_3 \times E_{h_3} & \\ & & \ddots \\ & & P_{p-1} \times E_{h_{p-1}} \end{array} \right\|$$

where P_i is a square matrix of h_i rows. In addition in order that P commute with $\mathfrak{M}$ the following relations must be satisfied

$$(14)\qquad \begin{aligned} A_1^2 P_1 &= P_2 A_1^2 \\ A_3^2 P_3 &= P_2 A_3^2 \\ &\cdots\cdots\cdots \\ A_{p-2}^{p-1} P_{p-2} &= P_{p-1} A_{p-2}^{p-1}. \end{aligned}$$

Here A_q^p has h_p rows, h_q columns.

We assume first that all h_i $(i = 1, 2, \cdots, p-1)$ are different from 0. Then the relations (14), together with the theorem[12] that a matrix P commuting with an indecomposable representation $\mathfrak{M}$ can have just one distinct characteristic root, are sufficient to show that each $h_i = 1$.

In outline the argument is this. We take P with $P_2 = P_3 = \cdots = P_{p-1} = 0$, then the only relation (14) remaining is

$$(15)\qquad A_1^2 P_1 = 0.$$

[12] Cf. Brauer-Schur, [11].

If now the rank of A_1^2 were less than h_1 we could find a P_1 with characteristic root $\neq 0$ to satisfy (15), contrary to the Schur-Brauer theorem. Then the rank of $A_1^2 = h_1$, and so $h_2 \geqq h_1$. Now taking $P_3 = P_4 = \cdots P_{p-1} = 0$, and choosing P_1 to satisfy $A_1^2 P_1 = P_2 A_1^2$, the remaining relation (14) is $P_2 A_3^2 = 0$ and by the same argument as before the rank of A_3^2 is h_2, and $h_3 \geqq h_2$. Next, we take $P_4 = P_5 = \cdots = P_{p-1} = 0$, choose P_1, P_2 to satisfy the first two relations (14), and consider $A_3^4 P_3 = 0$. We obtain in this manner that

$$h_1 \leqq h_2 \leqq h_3 \leqq \cdots \leqq h_{p-1}.$$

If, however, we had started at the other end of the relations (14), we would obtain

$$h_{p-1} \leqq h_{p-2} \leqq \cdots \leqq h_1$$

so that $h_1 = h_2 = \cdots = h_{p-1} = c$, say, and further the A_q^p are all of rank c and so are non-singular. It follows that the relations (14) are satisfied when P_1 is arbitrary, $P_2 = (A_1^2)^{-1} P_1 A_1^2$, $P_3 = (A_3^2)^{-1} P_2 A_3^2$, etc. Then P_1 must be of degree 1 (for otherwise P_1 could have two different characteristic roots) and so $c = 1$.

The second step in this argument requires some elaboration. Since we have seen A_1^2 is of rank h_1, we may find non-singular matrices X and Y such that $XA_1^2Y = \begin{Vmatrix} E_{h_1} \\ 0 \end{Vmatrix} = \bar{A}_1^2$, and setting $\bar{P}_1 = Y^{-1}P_1Y$, $\bar{P}_2 = XP_2X^{-1}$, we have from

$$\bar{A}_1^2 \bar{P}_1 = \bar{P}_2 \bar{A}_1^2 \tag{16}$$

that $\bar{P}_2$ has form

$$\bar{P}_2 = \begin{Vmatrix} P_1 & P_{12} \\ 0 & P_{22} \end{Vmatrix}.$$

We take $P_3 = P_4 = \cdots = P_{p-1} = 0$, assume that the rank of A_3^2 is less than h_2, and seek $\bar{P}_1$, $\bar{P}_2$ such that (16) is satisfied, and

$$\bar{P}_2 \bar{A}_3^2 = 0$$

where $\bar{A}_3^2 = XA_3^2$. Under our assumption that the rank of $A_3^2 < h_2$, we can find a vector $x = (0, 0, \cdots, 0, x_i, x_{i+1}, \cdots, x_{h_2})$, $x_i \neq 0$ which is annihilated by $\bar{A}_3^2$. Choose $\bar{P}_2$ to have the vector x in its i^{th} row, and zeros in the other rows. Then trace $\bar{P}_2 = x_i$, so that $\bar{P}_2$ has a characteristic root $\neq 0$. Further, for this choice of $\bar{P}_2$, we can always find $\bar{P}_1$ so that (16) is satisfied. Thus our assumption that A_3^2 has rank less than h_2 would lead to a matrix P which commutes with $\mathfrak{M}$ and has x_i and 0 for characteristic roots, which gives a contradiction.

For the case that not all $\mathfrak{F}_\kappa$ belonging to the block $\mathfrak{B}_r$ appear in the indecomposable representation study of the relations (14) show that the $\mathfrak{F}_\kappa$ which do appear have multiplicity 1 and form a sequence

$$\mathfrak{F}_\alpha, \mathfrak{F}_{\alpha+1}, \cdots, \mathfrak{F}_{\alpha+r}.$$

3) $l(\mathfrak{M}) = 3$. Here again the modular irreducible constituents of $\mathfrak{M}$ must belong to a block of lowest kind. We have also here that $\mathfrak{N}^2\mathfrak{B} \neq 0$. Then there is a Cartan basis element $b_{\kappa\kappa} \subset \mathfrak{N}^2$ such that $b_{\kappa\kappa}\mathfrak{B} \neq 0$, suppose that $b_{\kappa\kappa}x \neq 0$, $x \in \mathfrak{B}$. Since $b_{\kappa\kappa} = b_{\kappa,\kappa-1}b_{\kappa-1,\kappa}$ we have that $b_{\kappa-1,\kappa}x \neq 0$, similarly $b_{\kappa+1,\kappa}x \neq 0$, and from $b_{\kappa\kappa} = b_{\kappa\kappa}e_\kappa(11)$ we have also that $e_\kappa(11)x \neq 0$. Let, as above, $\mathfrak{l}_\kappa$ denote the indecomposable left ideal generated by the elements (10). Then it may be shown that $\mathfrak{l}_\kappa$ contains no left annihilators of x other than 0, so that

$$a \to ax, \qquad a \in \mathfrak{l}_\kappa$$

is an isomorphic mapping of $\mathfrak{l}_\kappa$ upon the invariant subspace $\mathfrak{B}_1 = \mathfrak{l}_\kappa x$ of $\mathfrak{B}$. It follows that $\mathfrak{M}$ contains a constituent equivalent to the indecomposable $\mathfrak{U}_\kappa$ of $\mathfrak{R}$ which corresponds to $\mathfrak{l}_\kappa$. This implies that the indecomposable representation $\mathfrak{M}$ is equal to $\mathfrak{U}_\kappa$.[13]

We gather these results in

THEOREM I. *An indecomposable representation $\mathfrak{M}$ of $\mathfrak{R}_m$ has Loewy length $l(\mathfrak{M}) \leqq 3$. If $l(\mathfrak{M}) = 1$, $\mathfrak{M}$ is equivalent to a modular irreducible part of $\mathfrak{R}_m$. For $l(\mathfrak{M}) = 2$, $\mathfrak{M}$ has the form* (12) *with $h_i = 1$ for a sequence of consecutive values of i, and the remaining $h_i = 0$, or $\mathfrak{M}$ is of the similar form but with the positions of the odd and the even constituents reversed. When $l(\mathfrak{M}) = 3$, $\mathfrak{M}$ is equivalent to an indecomposable part of the regular representation of $\mathfrak{R}_m$.*

6. Nakayama's results

We have seen above that the number of blocks of lowest kind is P_l, and that the number of blocks of highest kind is $P_m - pP_l$. Nakayama (8 II) has shown how to relate blocks and partitions more precisely. In the present section we state his results and give a slight extension of them.

Associated with the partition $(\lambda):\lambda_1 + \cdots + \lambda_k = m$, $(\lambda_1 \geqq \cdots \geqq \lambda_k > 0)$ of m is a diagram T consisting of k rows of *fields*, λ_i fields in the i^{th} row, the j^{th} elements of the rows being arranged in a column. The field in the i^{th} row and j^{th} column will be denoted by (i, j). By the (i, j)-*hook* $H = H(i, j)$ of T we mean the set of fields (i, v) with $v \geqq j$ together with the fields (v, j) with $v > i$. We call the number, h, of fields in H its *length*. If just r rows of T contain elements of H we call r the *height* of H. By $T - H$ we mean the diagram T' obtained from T by deleting the fields of H and then moving each field (i', j') with $i' > i, j' > j$ one row up and one column to the left.

We next divide the partions (λ) of m into classes according to the following rules. If the diagram, T, of (λ) has no hook of length p, then (λ) is put in a class by itself, called a *class of highest kind*. If T has a hook, H, of length p and height r, we denote (λ) by $\lambda^r(\mu)$ where (μ) is the partition whose diagram is $T - H$. *For $r = 1, \cdots, p$ there is exactly one partition $\lambda^r(\mu)$ for each partition (μ) of $l(= m - p)$.* The p partitions $\lambda^r(\mu)$ defined by (μ) are put into a class which we call a *class of lowest kind*. We can now state Nakayama's results.

[13] Cf. Remark 3, Nakayama-Nesbitt, [7].

Theorem II. *Let $\mathfrak{A}(\lambda)$ denote the ordinary irreducible representation of $\mathfrak{S}_m$ defined by (λ) and let $\mathfrak{F}(\lambda)$ be the modular representation induced by $\mathfrak{A}(\lambda)$. If (λ) belongs to a class of highest kind, then $\mathfrak{F}(\lambda)$ is irreducible and constitutes a block of highest kind. The p representations $\mathfrak{A}(\lambda^r(\mu))$ $r = 1, \cdots, p$, are the ordinary irreducible representations belonging to a block of lowest kind, which we accordingly denote by $\mathfrak{B}(\mu)$. We can enumerate the modular irreducible representations $\mathfrak{F}_1(\mu), \cdots, \mathfrak{F}_{p-1}(\mu)$ belonging to $\mathfrak{B}(\mu)$ so that*

$$\mathfrak{F}(\lambda^1(\mu)) \leftrightarrow \mathfrak{F}_1(\mu),\ \mathfrak{F}(\lambda^2(\mu)) \leftrightarrow \mathfrak{F}_1(\mu) + \mathfrak{F}_2(\mu), \cdots,$$

$$\mathfrak{F}(\lambda^{p-1}(\mu)) \leftrightarrow \mathfrak{F}_{p-2}(\mu) + \mathfrak{F}_{p-1}(\mu),\ \mathfrak{F}(\lambda^p(\mu)) \leftrightarrow \mathfrak{F}_{p-1}(\mu)$$

($\leftrightarrow$ denotes "has same irreducible constituents as" not "is equivalent to").

Let $(\mu) = (\mu_1, \cdots, \mu_k)$ (where $\mu_1 \geqq \mu_2 \geqq \cdots \geqq \mu_k > 0 = \mu_{k+1} = \cdots$) be a partition of l. If $\mu_i > \mu_{i+1}$ we denote by $(\mu \mid i)$ the partition $(\mu_1, \cdots, \mu_i - 1, \cdots, \mu_k)$ of $l - 1$. In the course of proving the above main theorem Nakayama showed that

$$(\lambda^j(\mu) \mid i) = \lambda^j(\mu \mid i) \tag{17}$$

except when (a) $i = 1$ and $\mu_j = \mu_{j+1}$ or (b) $i = j - 1$ and $\mu_j = \mu_{j-1}$. (In other words successive removal of hooks is almost a commutative process). This fact enables us to prove the following corollary[14] to the main theorem. If we consider a representation $\mathfrak{F}$ of $\mathfrak{S}_m$ only for permutations omitting the letter m we get a representation of $\mathfrak{S}_{m-1}$ which we denote by $\mathfrak{F}^*$.

Corollary I.

$$\begin{aligned}\mathfrak{F}_j(\mu)^* \leftrightarrow \textstyle\sum_i \mathfrak{F}_j(\mu \mid i) + \delta_{\mu_j\mu_{j+1}}\mathfrak{F}(\mu_j + p - j, \mu_1 + 1, \cdots,& \\ \mu_{j-1} + 1, \mu_{j+1}, \cdots, \mu_k)&\end{aligned} \tag{18}$$

The sum is over all i for which $\mu_i > \mu_{i+1}$.

Proof. It is well known[15] that $\mathfrak{A}(\lambda)^* \leftrightarrow \sum_i \mathfrak{A}(\lambda \mid i)$ where the sum extends over all i for which $\lambda_i > \lambda_{i+1}$. Applying Theorem II and equation (17) to the induced modular representation $\mathfrak{F}(\lambda)^*$ for $(\lambda) = \lambda^\rho(\mu)$ we get

$$\begin{aligned}\mathfrak{F}(\lambda^\rho(\mu))^* &\leftrightarrow \textstyle\sum_i \mathfrak{F}_{\rho-1}(\mu \mid i) + \mathfrak{F}_\rho(\mu \mid i) \\ &+ \delta_{\mu_\rho\mu_{\rho+1}}\mathfrak{F}(\lambda^\rho(\mu) \mid 1) + \delta_{\mu_{\rho-1}\mu_\rho}\mathfrak{F}(\lambda^\rho(\mu) \mid \rho - 1), \qquad p = 1, \cdots p.\end{aligned} \tag{19}$$

(The sum is over all i for which $\mu_i > \mu_{i+1}$.) We have also from the main theorem that

$$\mathfrak{F}^j(\mu) \leftrightarrow \mathfrak{F}(\lambda^j(\mu)) - \mathfrak{F}(\lambda^{j-1}(\mu)) \cdots (-1)^{j+1}\mathfrak{F}(\lambda^1(\mu)) \tag{20}$$

The corollary now follows at once by starring both sides of (20), substituting from (19) in each term on the right hand side and simplifying.

---

[14] Aside from this corollary everything in the present section is a restatement of Nakayama's results.

[15] [13] p. 215.

Observe that the terms on the right hand side of (18) all come from different blocks and so we can strengthen the corollary by the additional statement that *$\mathfrak{F}_i(\mu)^*$ is a completely reducible representation of $\mathfrak{S}_{m-1}$.*

7. Definitions and notation

In the next section we will consider the decomposition of representations of $\mathfrak{S}_m$ when considered as representations of various subgroups $\mathfrak{S}_{m-v}$. To facilitate this we introduce the following calculus of partitions.

Let (λ) be a partition of m (as in §6 above) and let $(i) = (i_1, \cdots, i_v)$ be a sequence consisting of μ_1 1's, μ_2 2's, $\cdots$, μ_k k's. By $(\lambda \mid i) = (\lambda \mid i_1 \cdots i_v)$ we mean the partition[16] $(\lambda_1 - \mu_1, \cdots, \lambda_k - \mu_k)$ of $m - v$. The sequence $i_1, \cdots, i_v$ is said to be (λ)-*proper* if for each ν from 1 to v the partition $(\lambda \mid i_1 \cdots i_\nu)$ of $m - \nu$ has non-negative terms in non-increasing order. Otherwise (i) is called (λ)-*improper*.

In this and the following sections we shall understand by $\mathfrak{A}(\lambda): s \rightarrow A(\lambda \between s)$ Young's rational semi-normal form[17] of the irreducible representations of $\mathfrak{S}_m$ defined by (λ).

If the sequence (i) is (λ)-proper we define $\mathfrak{A}(\lambda \mid i)$ to be the corresponding representation $s \rightarrow A(\lambda \mid i \between s)$ of $\mathfrak{S}_{m-v}$. If (i) is (λ)-improper we shall mean by $\mathfrak{A}(\lambda \mid i)$ a zero rowed square matrix. An important property of the semi-normal form is that $\mathfrak{A}(\lambda)$ when considered as a representation of $\mathfrak{S}_{m-v}$ (the symmetric group on the first $m - v$ letters of $\mathfrak{S}_m$) is already in completely reduced form with the $\mathfrak{A}(\lambda \mid i)$ as diagonal constituents. In other words each (λ)-proper sequence (i) of length v contributes an irreducible constituent. If (i) and (i') are two (λ)-proper sequences of length v, then $\mathfrak{A}(\lambda \mid i)$ appears above $\mathfrak{A}(\lambda \mid i')$ if the first nonvanishing difference $i_1 - i_1', \cdots, i_v - i_v'$ is positive.

We recall that the rows of $\mathfrak{A}(\lambda)$ are in 1-1 correspondence with the regular diagrams belonging to (λ). The representation $\mathfrak{A}(\lambda \mid i)$ occupies the (consecutive) rows whose corresponding diagrams have m in the i_1^{th} row, $m - 1$ in the i_2^{th} row, $\cdots$, $m - v + 1$ in the i_v^{th} row. By $A(\lambda \mid i \mid j \between s)$ we mean the rectangular submatrix of $A(\lambda \between s)$ whose rows are those of $\mathfrak{A}(\lambda \mid i)$ and whose columns are those of $\mathfrak{A}(\lambda \mid j)$. By the v^{th} *refinement of* $\mathfrak{A}(\lambda)$ we mean that each $A(\lambda \between s)$ is to be considered as a matrix whose elements are the submatrices $A(\lambda \mid i \mid j \between s)$.

8. The representations of $\mathfrak{S}_p$

For $m = p$ every $A(\lambda \between s)$ is p-integral. There is just one block $\mathfrak{B} = \mathfrak{B}(0)$ of lowest kind. From Theorem II it follows that $\mathfrak{A}(\lambda)$ belongs to $\mathfrak{B}$ if and only if the diagram of (λ) is a hook, i.e. one of the partitions $(p - \rho, 1^\rho)$, $\rho = 0, \cdots, p - 1$.

---

[16] Note that $\lambda_i \geqq \mu_i$ is not required since for some purposes partitions with negative summands are useful.

[17] [14] pp. 217–88 or [12].

Let $(\lambda) = (p - \rho, 1^\rho)$ where $0 < \rho < p - 1$. There are just four (λ)-proper sequences of length 2. These are (in order) $(i^1) = (\rho + 1, \rho)$, $(i^2) = (\rho + 1, 1)$, $(i^3) = (1, \rho + 1)$, $(i^4) = (1, 1)$. Note that the diagram of $(\lambda \mid i^r)$ is a hook of length $p - 2$. We denote by (λ^ρ) the partition $(p - \rho - 1, 1^{\rho-1})$ of $p - 2$, and by g^ρ the degree of the representation $\mathfrak{A}(\lambda^\rho)$ of $\mathfrak{S}_{p-2}$. In this notation we have $\mathfrak{A}(\lambda \mid i^1) = \mathfrak{A}(\lambda^{\rho-1})$, $\mathfrak{A}(\lambda \mid i^2) = \mathfrak{A}(\lambda \mid i^3) = \mathfrak{A}(\lambda^\rho)$, and $\mathfrak{A}(\lambda \mid i^4) = \mathfrak{A}(\lambda^{\rho+1})$.

For elements of $\mathfrak{S}_{p-2}$ the non diagonal parts of the second refinement of $\mathfrak{A}(\lambda)$ all vanish and along the main diagonal we have $\mathfrak{A}(\lambda^{\rho-1})$, $\mathfrak{A}(\lambda^\rho)$, $\mathfrak{A}(\lambda^\rho)$, $\mathfrak{A}(\lambda^{\rho+1})$ in the order named. Let t_r denote the transposition $(r - 1, r)$. Then $A(\lambda \mid i^j \mid i^k \between t_{p-1}) = 0$ for $j \neq k$ unless $(j, k) = (1, 2), (2, 1), (3, 4)$ or $(4, 3)$; and $A(\lambda \mid i^j \mid i^k \between t_p) = 0$ for $j \neq k$ unless $(j, k) = (2, 3)$ or $(3, 2)$, the non zero parts being scalar; say $A(\lambda \mid i^j \mid i^k \between t_p) = \alpha_{jk} E_{jk}$ where E_{jk} is the identity matrix of suitable degree. For the α_{jk} we have $\alpha_{11} = +1$, $\alpha_{22} = -1/(p - 1)$, $\alpha_{23} = (p^2 - 2p)/(p - 1)^2$, $\alpha_{32} = 1$, $\alpha_{33} = 1/(p - 1)$, $\alpha_{44} = 1$.

Now consider the induced modular representation $\mathfrak{F}(\lambda)$. Since $\mathfrak{A}(\lambda)$ is p-integral we are justified in keeping the same refinement notation, i.e. we write $F(\lambda \between s) = \| F(\lambda \mid i^j \mid i^k \between s) \|$ where each submatrix of $F(\lambda \between s)$ is obtained by considering the corresponding submatrix of $A(\lambda \between s)$ mod p. For $s \in \mathfrak{S}_{p-1}$ we have

$$F(\lambda \between s) = \left\| \begin{array}{cc} F(\lambda \mid \rho + 1 \between s) & \begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \\ \begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} & F(\lambda \mid 1 \between s) \end{array} \right\|$$

where $F(\lambda \mid \rho + 1 \between s)$ occupies the first two sets of rows and columns and $F(\lambda \mid 1 \between s)$ the last two sets. $F(\lambda \mid \rho + 1)$ is (as the notation implies) $A(\lambda \mid \rho + 1)$ taken mod p.

Furthermore, we have

$$F(\lambda \between t_p) = \left\| \begin{array}{cccc} -E_{11} & 0 & 0 & 0 \\ 0 & E_{22} & 0 & 0 \\ 0 & E_{32} & -E_{33} & 0 \\ 0 & 0 & 0 & E_{44} \end{array} \right\|.$$

Since $\mathfrak{S}_p$ is generated by $\mathfrak{S}_{p-1}$ and t_p, the forms of these matrices show just how to write down the irreducible constituents $(\mathfrak{F}_\rho(0), \mathfrak{F}_{\rho+1}(0))$ of $\mathfrak{F}(\lambda)$. (For $\rho = 0$ and $\rho = p - 1$ the situation differs only in that certain indicated parts of the refinement do not appear, and of course there are no representations $\mathfrak{F}_\rho(-1)$ and $\mathfrak{F}_\rho(p)$). Summarizing we see that $\mathfrak{F}_\rho(0 \between s) = \mathfrak{F}(p - \rho, 1^{\rho-1} \between s)$ if $s \in \mathfrak{S}_{p-1}$ and $F_\rho(0 \between t_p) = \left\| \begin{array}{cc} -E_{11} & 0 \\ 0 & E_{22} \end{array} \right\|$ where E_{11} is of degree $g^{\rho-1}$ and E_{22} of degree g^ρ.

This completes the determination of the irreducible representations of $\mathfrak{S}_p$.

But we can get still more information from the above process. For by Theorem I it follows that the lower left hand corner (last two sets of rows first two sets of columns) of $\mathfrak{F}(\lambda)$ must be a numerical multiple of $\mathfrak{H}_\rho^{\rho+1}(0)$. We may, and do, suppose that the Cartan basis for $\mathfrak{R}_p$ was so chosen that this numerical multiple (which is obviously not zero) is unity. Then we have $H_\rho^{\rho+1}(0 \between s) = 0$ if $s \in \mathfrak{S}_{p-1}$ and $H_\rho^{\rho+1}(0 \between t_p) = \begin{Vmatrix} 0 & E_{32} \\ 0 & 0 \end{Vmatrix}$ (E_{32} of degree g^ρ).

To calculate $\mathfrak{H}_{\rho+1}^\rho(0)$ we replace the above used form of Young's semi-normal representation by the one generated by the transposes of the matrices $A(\lambda \between t_r)$ $r = 2, \cdots, p$; and we get $H_{\rho+1}^\rho(0 \between t_p) = \begin{Vmatrix} 0 & 0 \\ E_{23} & 0 \end{Vmatrix}$ (i.e. the transpose of $H_\rho^{\rho+1}(0 \between t_p)$), and $H_{\rho+1}^\rho(0 \between s) = 0$ if $s \in S_{p-1}$.

The final step in our program of determining all representations of $\mathfrak{S}_p$ is the calculation for the elementary modules $\mathfrak{H}_\rho^\rho(0)$ of unmixed type. This calculation is based upon the form (formula (11) above) of $\mathfrak{U}_\rho(0)$ and the following two facts: (1) $t_r^2 = 1$ and (2) t_p commutes with every element of $\mathfrak{S}_{p-2}$. It follows at once from $t_r^2 = 1$, $H_\rho^{\rho+1}(0 \between t_r) = H_{\rho+1}^\rho(0 \between t_r) = 0$ that $H_\rho^\rho(0 \between t_r) = 0$ for $r < p$. Since $\mathfrak{S}_{p-1}$ is generated by $t_2, \cdots, t_{p-1}$ this shows that $H_\rho^\rho(0 \between s) = 0$ if $s \in \mathfrak{S}_{p-1}$. It remains therefore only to calculate $H_\rho^\rho(0 \between t_p)$.

Considered as a representation of $\mathfrak{S}_{p-2}$, $\mathfrak{U}_\rho(0)$ takes the form

$$\begin{Vmatrix} \mathfrak{F}(\lambda^{\rho-1}) & & & & & & \\ & \mathfrak{F}(\lambda^{\rho}) & & & & & \\ & & \mathfrak{F}(\lambda^{\rho-2}) & & & & \\ & & & \mathfrak{F}(\lambda^{\rho-1}) & & & \\ & & & & \mathfrak{F}(\lambda^{\rho}) & & \\ & & & & & \mathfrak{F}(\lambda^{\rho-1}) & \\ & & & & & & \mathfrak{F}(\lambda^{\rho}) \end{Vmatrix}$$

Since t_p commutes with $\mathfrak{R}_{p-2}$ it now follows from Schur's Lemma that $H_\rho^\rho(0 \between t_p) = \begin{Vmatrix} \alpha_1 E_{11} & 0 \\ 0 & \alpha_2 E_{22} \end{Vmatrix}$, ($E_{11}$ and E_{22} as above). Now apply the condition $t_p^2 = 1$ and we get $\alpha_1 = -\alpha_2 = 1/2$.

9. Further specific results

The above methods can be applied without serious additional complications (save in notation) to obtain similar results for $\mathfrak{S}_{p+1}$ and $\mathfrak{S}_{p+2}$. For $\mathfrak{S}_{p+1}$ there is again just one block of lowest kind and the ordinary representations belonging to it are p-integral (i.e. when put in rational semi-normal form). For $\mathfrak{S}_{p+2}$ there are two blocks of lowest kind, and although some of the ordinary semi-normal representations belonging to these blocks fail to be p-integral, they become so after a simple transformation which does not irreparably upset the refinement

Let $(\lambda) = (p - \rho, 1^\rho)$ where $0 < \rho < p - 1$. There are just four (λ)-proper sequences of length 2. These are (in order) $(i^1) = (\rho + 1, \rho)$, $(i^2) = (\rho + 1, 1)$, $(i^3) = (1, \rho + 1)$, $(i^4) = (1, 1)$. Note that the diagram of $(\lambda \mid i^\nu)$ is a hook of length $p - 2$. We denote by (λ^ρ) the partition $(p - \rho - 1, 1^{\rho-1})$ of $p - 2$, and by g^ρ the degree of the representation $\mathfrak{A}(\lambda^\rho)$ of $\mathfrak{S}_{p-2}$. In this notation we have $\mathfrak{A}(\lambda \mid i^1) = \mathfrak{A}(\lambda^{\rho-1})$, $\mathfrak{A}(\lambda \mid i^2) = \mathfrak{A}(\lambda \mid i^3) = \mathfrak{A}(\lambda^\rho)$, and $\mathfrak{A}(\lambda \mid i^4) = \mathfrak{A}(\lambda^{\rho+1})$.

For elements of $\mathfrak{S}_{p-2}$ the non diagonal parts of the second refinement of $\mathfrak{A}(\lambda)$ all vanish and along the main diagonal we have $\mathfrak{A}(\lambda^{\rho-1})$, $\mathfrak{A}(\lambda^\rho)$, $\mathfrak{A}(\lambda^\rho)$, $\mathfrak{A}(\lambda^{\rho+1})$ in the order named. Let t_r denote the transposition $(r - 1, r)$. Then $A(\lambda \mid i^j \mid i^k \between t_{p-1}) = 0$ for $j \neq k$ unless $(j, k) = (1, 2), (2, 1), (3, 4)$ or $(4, 3)$; and $A(\lambda \mid i^j \mid i^k \between t_p) = 0$ for $j \neq k$ unless $(j, k) = (2, 3)$ or $(3, 2)$, the non zero parts being scalar; say $A(\lambda \mid i^j \mid i^k \between t_p) = \alpha_{jk} E_{jk}$ where E_{jk} is the identity matrix of suitable degree. For the α_{jk} we have $\alpha_{11} = +1$, $\alpha_{22} = -1/(p - 1)$, $\alpha_{23} = (p^2 - 2p)/(p - 1)^2$, $\alpha_{32} = 1$, $\alpha_{33} = 1/(p - 1)$, $\alpha_{44} = 1$.

Now consider the induced modular representation $\mathfrak{F}(\lambda)$. Since $\mathfrak{A}(\lambda)$ is p-integral we are justified in keeping the same refinement notation, i.e. we write $F(\lambda \between s) = \| F(\lambda \mid i^j \mid i^k \between s) \|$ where each submatrix of $F(\lambda \between s)$ is obtained by considering the corresponding submatrix of $A(\lambda \between s)$ mod p. For $s \in \mathfrak{S}_{p-1}$ we have

$$F(\lambda \between s) = \left\| \begin{array}{cc} F(\lambda \mid \rho + 1 \between s) & \begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \\ \begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} & F(\lambda \mid 1 \between s) \end{array} \right\|$$

where $F(\lambda \mid \rho + 1 \between s)$ occupies the first two sets of rows and columns and $F(\lambda \mid 1 \between s)$ the last two sets. $F(\lambda \mid \rho + 1)$ is (as the notation implies) $A(\lambda \mid \rho + 1)$ taken mod p.

Furthermore, we have

$$F(\lambda \between t_p) = \left\| \begin{array}{cccc} -E_{11} & 0 & 0 & 0 \\ 0 & E_{22} & 0 & 0 \\ 0 & E_{32} & -E_{33} & 0 \\ 0 & 0 & 0 & E_{44} \end{array} \right\|.$$

Since $\mathfrak{S}_p$ is generated by $\mathfrak{S}_{p-1}$ and t_p, the forms of these matrices show just how to write down the irreducible constituents $(\mathfrak{F}_\rho(0), \mathfrak{F}_{\rho+1}(0))$ of $\mathfrak{F}(\lambda)$. (For $\rho = 0$ and $\rho = p - 1$ the situation differs only in that certain indicated parts of the refinement do not appear, and of course there are no representations $\mathfrak{F}_\rho(-1)$ and $\mathfrak{F}_\rho(p)$). Summarizing we see that $\mathfrak{F}_\rho(0 \between s) = \mathfrak{F}(p - \rho, 1^{\rho-1} \between s)$ if $s \in \mathfrak{S}_{p-1}$ and $F_\rho(0 \between t_p) = \left\| \begin{array}{cc} -E_{11} & 0 \\ 0 & E_{22} \end{array} \right\|$ where E_{11} is of degree $g^{\rho-1}$ and E_{22} of degree g^ρ.

This completes the determination of the irreducible representations of $\mathfrak{S}_p$.

But we can get still more information from the above process. For by Theorem I it follows that the lower left hand corner (last two sets of rows first two sets of columns) of $\mathfrak{F}(\lambda)$ must be a numerical multiple of $\mathfrak{H}_\rho^{\rho+1}(0)$. We may, and do, suppose that the Cartan basis for $\mathfrak{R}_p$ was so chosen that this numerical multiple (which is obviously not zero) is unity. Then we have $H_\rho^{\rho+1}(0)(s) = 0$ if $s \in \mathfrak{S}_{p-1}$ and $H_\rho^{\rho+1}(0)(t_p) = \left\| \begin{matrix} 0 & E_{32} \\ 0 & 0 \end{matrix} \right\|$ (E_{32} of degree g^ρ).

To calculate $\mathfrak{H}_{\rho+1}^\rho(0)$ we replace the above used form of Young's semi-normal representation by the one generated by the transposes of the matrices $A(\lambda)(t_r)$ $r = 2, \cdots, p$; and we get $H_{\rho+1}^\rho(0)(t_p) = \left\| \begin{matrix} 0 & 0 \\ E_{23} & 0 \end{matrix} \right\|$ (i.e. the transpose of $H_\rho^{\rho+1}(0)(t_p)$), and $H_{\rho+1}^\rho(0)(s) = 0$ if $s \in S_{p-1}$.

The final step in our program of determining all representations of $\mathfrak{S}_p$ is the calculation for the elementary modules $\mathfrak{H}_\rho^\rho(0)$ of unmixed type. This calculation is based upon the form (formula (11) above) of $\mathfrak{U}_\rho(0)$ and the following two facts: (1) $t_r^2 = 1$ and (2) t_p commutes with every element of $\mathfrak{S}_{p-2}$. It follows at once from $t_r^2 = 1$, $H_\rho^{\rho+1}(0)(t_r) = H_{\rho+1}^\rho(0)(t_r) = 0$ that $H_\rho^\rho(0)(t_r) = 0$ for $r < p$. Since $\mathfrak{S}_{p-1}$ is generated by $t_2, \cdots, t_{p-1}$ this shows that $H_\rho^\rho(0)(s) = 0$ if $s \in \mathfrak{S}_{p-1}$. It remains therefore only to calculate $H_\rho^\rho(0)(t_p)$.

Considered as a representation of $\mathfrak{S}_{p-2}$, $\mathfrak{U}_\rho(0)$ takes the form

$$\left\| \begin{matrix} \mathfrak{F}(\lambda^{\rho-1}) & & & & & & \\ & \mathfrak{F}(\lambda^{\rho}) & & & & & \\ & & \mathfrak{F}(\lambda^{\rho-2}) & & & & \\ & & & \mathfrak{F}(\lambda^{\rho-1}) & & & \\ & & & & \mathfrak{F}(\lambda^{\rho}) & & \\ & & & & & \mathfrak{F}(\lambda^{\rho-1}) & \\ & & & & & & \mathfrak{F}(\lambda^{\rho}) \end{matrix} \right\|$$

Since t_p commutes with $\mathfrak{R}_{p-2}$ it now follows from Schur's Lemma that $H_\rho^\rho(0)(t_p) = \left\| \begin{matrix} \alpha_1 E_{11} & 0 \\ 0 & \alpha_2 E_{22} \end{matrix} \right\|$, ($E_{11}$ and E_{22} as above). Now apply the condition $t_p^2 = 1$ and we get $\alpha_1 = -\alpha_2 = 1/2$.

9. Further specific results

The above methods can be applied without serious additional complications (save in notation) to obtain similar results for $\mathfrak{S}_{p+1}$ and $\mathfrak{S}_{p+2}$. For $\mathfrak{S}_{p+1}$ there is again just one block of lowest kind and the ordinary representations belonging to it are p-integral (i.e. when put in rational semi-normal form). For $\mathfrak{S}_{p+2}$ there are two blocks of lowest kind, and although some of the ordinary semi-normal representations belonging to these blocks fail to be p-integral, they become so after a simple transformation which does not irreparably upset the refinement

into submatrices. But for $\mathfrak{S}_{p+l}$ with $l > 2$ no such simple transformation into p-integral form has yet been found.

One might hope to get further information by starting with one of the known integral forms for the irreducible representations of $\mathfrak{S}_m$, rather than with Young's semi-normal form. The drawback to such a procedure is that these integral forms are not adapted for descent to the subgroups $\mathfrak{S}_{m-v}$ of $\mathfrak{S}_m$. So when they are taken mod p their decomposition seems to be almost (if not fully) as difficult as that of the regular representation, and so there is no particular point in using them.

INSTITUTE FOR ADVANCED STUDY AND
UNIVERSITY OF MICHIGAN

BIBLIOGRAPHY

1. R. BRAUER AND C. NESBITT, *On the Modular Characters of Groups*, these Annals, (2), vol. 42, (1941), pp. 556–590.
2. R. BRAUER, *Investigations on Group Characters*, these Annals, (4), vol. 42, (1941), pp. 936–958.
3. R. BRAUER, *On the Cartan Invariants of Groups of Finite Order*, these Annals, (1), vol. 42, (1941), pp. 53–61.
4. R. BRAUER, *On Sets of Matrices with Coefficients in a Division Ring*, Trans. Amer. Math. Soc. (3), vol. 49, (1941), pp. 502–548.
5. R. BRAUER, *On the Representation of Groups of Finite Order*, Proceedings of the National Academy of Sciences, vol. 25, (1939), pp. 290–295.
6. C. NESBITT, *On the Regular Representations of Algebras*, these Annals, (3), vol. 39, (1938), pp. 634–658.
7. T. NAKAYAMA AND C. NESBITT, *Note on Symmetric Algebras*, these Annals, (3), vol. 39, (1938), pp. 659–668.
8. T. NAKAYAMA, *On Some Modular Properties of Irreducible Representations of a Symmetric Group I*, Japanese Journal of Math., vol. XVII, (1940), pp. 89–108; *II*, Japanese Journal of Math., vol. XVII, (1941), 411–423.
9. T. NAKAYAMA, *Some Studies on Regular Representations and Modular Representations*, these Annals, (2), vol. 39, (1938), pp. 361–369.
10. W. M. SCOTT, *On Matrix Algebras Over an Algebraically Closed Field*, these Annals (2), vol. 43, (1942), pp. 147–160.
11. R. BRAUER AND I. SCHUR, *Zum Irreduzibilitätsbegriff in der Gruppen linearer homogener Substitutionen*, Sitzungsber. Preuss. Akad., 1930, p. 209, §2.
12. R. M. THRALL, *On Young's Semi-normal Representation of the Symmetric Group*, Duke Journal, vol. 8, (1941), pp. 611–624.
13. H. WEYL, *The Classical Groups*, Princeton, 1939.
14. A. YOUNG, *Quantitative Substitutional Analysis, Part VI*, Proceedings of the London Mathematical Society (2), vol. 34, (1932), pp. 196–230.

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

ON THE DECOMPOSITION OF MODULAR TENSORS (I)

BY R. M. THRALL

(Received December 11, 1941)

1. Introduction

This paper is a companion to the preceding one[1] [5], so we number formulas, theorems, etc., consecutively from those in that paper, and preserve the same notation unless a change is specifically indicated.

Let $\mathfrak{V}_1$ be a vector space over a field $\mathfrak{k}$ of characteristic p. We are interested in the decomposition of the space $\mathfrak{V}_m$ of all tensors of rank m, relative to the Kronecker m^{th} power representation Π_m of the group $\mathfrak{G}$ of all non-singular linear transformations of $\mathfrak{V}_1$ into itself. In this paper we determine the structure of $\mathfrak{V}_m$ subject to the two limitations: I. $m < 2p$; II. $\mathfrak{k}$ has at least m elements. The first limitation is due to the incomplete state of the theory of modular representations of the symmetric group. The second limitation is less serious, although the decomposition is actually different if $\mathfrak{k}$ has less than m elements. We hope to treat this case in a later paper.

The principal results about $\mathfrak{V}_m$ are contained in Theorems III and VII, together with formula (38). The problem is attacked by exhibiting the enveloping algebra of Π_m as the commutator algebra of a certain permutation representation of the symmetric group of degree m. A main tool in this process is application of Remark I, below, which states that the order of the commutator algebra of a group of permutation matrices is independent of the underlying field, i.e. is even the same characteristic 0 as characteristic p.

2. The commutator algebra of a monomial group

An n-rowed matrix is called *monomial* if it has exactly one non-zero element in each row and column. A *permutation matrix* is a monomial matrix in which each of the n non-zero elements is 1. A *diagonal matrix* is a monomial matrix whose non-zero terms all lie on the main diagonal.

Let $\mathfrak{A}$: $s \rightarrow A(s) = \| a_{ij}(s) \|$ be a monomial $\mathfrak{k}$-representation of degree n of a group $\mathfrak{G}$; $\mathfrak{k}$ being any field. (We call $\mathfrak{A}$ monomial when each $A(s)$ is monomial.) The set $\mathfrak{B}$ of all $\mathfrak{k}$-matrices B such that

$$A(s)B = BA(s) \qquad \text{for all } s \text{ in } \mathfrak{G} \tag{21}$$

is called the *commutator algebra* of the representation $\mathfrak{A}$. We are interested in determining the nature and order of $\mathfrak{B}$.

We first treat the case in which $\mathfrak{A}$ is a permutation representation. Suppose that

[1] Brackets refer to the bibliography at the end of the paper.

$$A(s)\begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix} = \begin{Vmatrix} x_{1(s)} \\ \vdots \\ x_{n(s)} \end{Vmatrix} \tag{22}$$

then (21) in the form $A(s)BA(s)^{-1} = B$ is equivalent to the equations

$$b_{i(s)\,j(s)} = b_{ij} \quad i, j = 1, \cdots, n, \text{ for all } s \text{ in } \mathfrak{G}. \tag{23}$$

We divide the index pairs (i, j) into systems of transitivity according to the equivalence relation:

$$(i, j) \sim (i', j') \leftrightarrow \text{there is an } s \text{ in } \mathfrak{G} \text{ for which } i' = i(s);\ j' = j(s). \tag{24}$$

It is clear from (23) that B is a commutator of $\mathfrak{A}$ if and only if

$$b_{ij} = b_{i'j'} \text{ whenever } (i, j) \sim (i', j'). \tag{25}$$

LEMMA I. *The order of the commutator algebra of a permutation representation (of a finite group) is equal to the number of systems of transitivity in the Kronecker square of the representation.*

PROOF. We can consider the n^2 numbers $x_i y_j$ as coordinates of the general vector in the space on which $\mathfrak{A} \times \mathfrak{A}$ operates. Then it follows from (22) that $A(s) \times A(s)$ sends the column matrix $\| x_i y_j \|$ into the column matrix $\| x_{i(s)} y_{j(s)} \|$. Thus the systems of transitivity under $\mathfrak{A} \times \mathfrak{A}$ are just the same as the systems of transitivity of the index pairs (i, j) defined by (24) above. With a system of transitivity we associate the matrix B having $b_{ij} = 1$ if (i, j) is in the given system and $b_{ij} = 0$ otherwise. The matrices thus constructed are clearly linearly independent; and it follows from (25) that they constitute a basis for $\mathfrak{B}$.

A monomial matrix A can be represented uniquely as the product DP of a diagonal matrix and a permutation matrix. If $A' = D'P'$ is a second monomial matrix, then the product $A'' = AA' = D''P''$ has $D'' = DPD'P^{-1}$ and $P'' = PP'$. Returning now to the arbitrary monomial representation $\mathfrak{A}$ we write $A(s) = D(s)P(s)$; $D(s) = \| \delta_{ij} d_i(s) \|$. We have just proved that $P(st) = P(s)P(t)$ and so $s \to P(s)$ is a permutation representation $\mathfrak{P}$ of $\mathfrak{G}$ which we call *the permutation representation belonging to* $\mathfrak{A}$.

The equations (21) are now equivalent to

$$d_i(s) b_{i(s)\,j(s)} / d_j(s) = b_{ij} \qquad \text{for all } s \text{ in } \mathfrak{G},\ i, j = 1, \cdots, m. \tag{26}$$

We divide the index pairs (i, j) into systems of transitivity according to $\mathfrak{P}$. Consider the subgroup $\mathfrak{H} = \mathfrak{H}(i, j)$ containing all s for which $i = i(s)$, $j = j(s)$. We say that the system of transitivity containing (i, j) is *singular* if for some s in $\mathfrak{H}$, $d_i(s) \neq d_j(s)$. [A simple computation shows that being singular is a property of the system of transitivity, independent of the representative (i, j) used to test for singularity.] If (i, j) belongs to a singular system we have $b_{ij} = 0$ by (26) and then by transitivity $b_{i'j'} = 0$ for every $(i', j') \sim (i, j)$. However, *if (i, j) belongs to a non-singular system then there is a commutator B*

with $b_{ij} = 1$, *and* $b_{i'j'} \neq 0$ *if and only if* $(i', j') \sim (i, j)$. The equations (26) will never lead to relations connecting b_{ij} and $b_{i'j'}$ unless $(i, j) \sim (i', j')$, so if there is any solution with $b_{ij} = 1$, there is one with $b_{ij} = 1$ and $b_{i'j'} = 0$ if (i', j') is not in the same system as (i, j). Since the $A(s)$ form a group, any equation in (26) connecting $b_{i(s)j(s)}$ and $b_{i(t)j(t)}$ is implied by those connecting b_{ij} to $b_{i'j'}$ with $(i', j') = (i(s), j(s)), (i(t), j(t)), (i(ts^{-1}), j(ts^{-1}))$. Hence our problem is reduced to showing that the equations (26) with b_{ij} on the right-hand side, are soluble with $b_{ij} = 1$. We know that for s in $\mathfrak{H}$ they are consistent. If $(i(s), j(s)) = (i(t), j(t))$ then $u = ts^{-1} \epsilon \mathfrak{H}$ and now calculating $d_i(t), d_j(t)$ from the equation $A(t) = A(u)A(s)$, we get $d_i(t)/d_j(t) = d_i(s)/d_j(s)$ which establishes the consistency of the equations. We have now proved

LEMMA II. *The order of the commutator algebra of a group of monomial matrices is equal to the number of non-singular transitive systems of index pairs.*

Let $\mathfrak{k}$ and $\mathfrak{K}$ be any two fields. A group of permutation matrices can be regarded as lying in either $\mathfrak{k}$ or $\mathfrak{K}$, since 1 is an element of any field. Lemma I states that the order of the $\mathfrak{k}$-commutator algebra is equal to the order of the $\mathfrak{K}$-commutator algebra. We shall apply this fact below to the case where one field is of characteristic 0 and the other of characteristic p. For future reference we restate this (weaker) form of Lemma I as

REMARK I. *The order of the commutator algebra of a group of permutation matrices is independent of the field of coefficients.*

The analogue to Remark I for monomial groups is not true. For consider the group of order p generated by $A(s) = \begin{Vmatrix} w & 0 \\ 0 & 1 \end{Vmatrix}$ where w is a p^{th} root of unity (in a field of characteristic 0). Since $w \equiv 1 \bmod p$, the modular image of this group is generated by $A(s) = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}$. The non-modular commutator algebra has order 2 and the modular one order 4.

3. Vanishing forms

Suppose that $\mathfrak{k}$ is the Galois field with q elements. Let $Y_1, Y_2, \cdots$ be indeterminants over $\mathfrak{k}$, and let $y_1, y_2, \cdots$ be variables whose domain is $\mathfrak{k}$. Consider the ideal in $\mathfrak{k}[Y_1, Y_2, \cdots]$ generated by $Y_i^q - Y_i$, $i = 1, 2, \cdots$. Modulo this ideal every $\mathfrak{k}$-polynomial $F(Y_1, Y_2, \cdots)$ is congruent to a unique $\mathfrak{k}$-polynomial $F^*(Y_1, Y_2, \cdots)$ of degree less than q in each Y_i, and $F(y_1, y_2, \cdots) = 0$ is equivalent to $F^*(Y_1, Y_2, \cdots) = 0$. In particular, a $\mathfrak{k}$-polynomial $F(Y_1)$ of degree less than q cannot vanish over $\mathfrak{k}$ (i.e. $F(y_1) = 0$) unless every coefficient is zero.

LEMMA III. *There is no non-zero* $\mathfrak{k}$*-form* $P(Y_{ij})$ *of degree* $m \leqq q$ *in the* n^2 *indeterminants* Y_{ij}, *such that* $P(y_{ij}) \det \| y_{ij} \| = 0$ *where the* y_{ij} *are* n^2 *variables whose domain is* $\mathfrak{k}$.

PROOF. If $m < q - 1$, then $F(Y_{ij}) = P(Y_{ij}) \det \| Y_{ij} \|$ is of degree less than q in each Y_{ij}, so $P(Y_{ij}) \neq 0$ implies $F(Y_{ij}) = F^*(Y_{ij}) \neq 0$ and therefore $F(y_{ij}) \neq 0$. This leaves the two cases $m = q - 1$, $m = q$. The proofs for

these are similar, so we treat here only the harder case, $m = q$. Write $P(Y_{ij})$ as a polynomial in Y_{11} with coefficients $P_i = P_i(Y_{ij})$ that are polynomials in $\mathfrak{k}[Y_{12}, \cdots, Y_{nn}]$, i.e.

$$P(Y_{ij}) = P_0 Y_{11}^m + P_1 Y_{11}^{m-1} + \cdots + P_m .$$

We also write

$$\det \| Y_{ij} \| = Q_0 Y_{11} + Q_1 .$$

Then

$$F(Y_{ij}) = Q_0 P_0 Y_{11}^{m+1} + (Q_0 P_1 + Q_1 P_0) Y_{11}^m + \cdots + Q_1 P_m .$$

If we replace Y^{m+1} by Y^2 and Y^m by Y, $F(Y_{ij})$ is replaced by the congruent polynomial (no longer a form)

$$F'(Y_{ij}) = (Q_0 P_2 + Q_1 P_1) Y_{11}^{m-1} + \cdots + (Q_0 P_{m-2} + Q_1 P_{m-3}) Y_{11}^3$$
$$+ (Q_0 P_0 + Q_0 P_{m-1} + Q_1 P_{m-2}) Y_{11}^2 + (Q_0 P_1 + Q_1 P_0 + Q_1 P_{m-1}) Y_{11} + Q_1 P_m .$$

Since $F'(Y_{ij})$ is of degree $q - 1$ (or less) in Y_{11} we cannot have $F'(y_{ij}) = 0$ unless the coefficient of each power of Y_{11} becomes zero when Y_{ij} is replaced by y_{ij}. For $i > 2$ the coefficient of Y_{11}^i is of degree less than q in each indeterminant and so vanishing for y_{ij} is the same as vanishing for Y_{ij}; i.e.

$$Q_0 P_i = -Q_1 P_{i-1} \qquad i = 2, \cdots, m - 2. \tag{27}$$

The lemma is trivial if $n = 1$. For $n > 1$, Q_0 and Q_1 are relatively prime forms of degree > 1. (27) for $i = 2$ requires $P_1 = P_2 = 0$, for otherwise P_1, of degree 1, would be divisible by Q_0, of degree > 1. Hence, $P_1 = P_2 = \cdots = P_{m-2} = 0$. We can apply the same process to each Y_{ij} and conclude that $F(y_{ij}) = 0$ *implies that* $P(Y_{ij})$ *has no terms of degree different from* m, 1, 0 *in each* Y_{ij}. But then $F^*(Y_{ij}) = P^*(Y_{ij}) \det \| Y_{ij} \|$ and $P^*(Y_{ij})$ is clearly not zero unless $P(Y_{ij}) = 0$; this completes the proof for $m = q$. Observe that the form $P(Y) = Y_{11}^q Y_{12} - Y_{11} Y_{12}^q$ shows that the lemma is false for $m > q$.

We use Lemma III as a modular substitute for what H. Weyl [6, p. 4] calls the "principle of the irrelevance of algebraic inequalities."

4. General program

Henceforth $\mathfrak{k}$ shall be a field of characteristic p. We are interested in the decomposition of the Kronecker m^{th} power representation $\Pi_m : A \rightarrow \Pi_m(A) = A \times \cdots \times A$ (m-factors) of the full linear group $\mathfrak{G} = \mathfrak{G}(n, \mathfrak{k})$ of n-rowed non-singular $\mathfrak{k}$-matrices. We propose to effect this decomposition by exhibiting the enveloping algebra $\mathfrak{A}_m$ of Π_m as the commutator algebra of a certain permutation representation of the symmetric group $\mathfrak{S}_m$ of degree m.

The order of $\mathfrak{A}_m$ is the number of linearly independent monomials of degree m in $N = n^2$ variables a_{ij} which range freely over $\mathfrak{k}$ save for the restriction $\det \| a_{ij} \| \neq 0$. Hence, by Lemma III we have

LEMMA IV. *If $\mathfrak{k}$ has at least m elements, then the order of $\mathfrak{A}_m$ is* $\binom{N+m-1}{m}$; *i.e. is equal to the number of linearly independent monomials of degree m in N indeterminants.*

Let $x(i_1, \cdots, i_m)(i_1, \cdots, i_m = 1, \cdots, n)$ be the components of an arbitrary vector in the $\mathfrak{k}$-space $\mathfrak{V}_m$ of all tensors of rank m; i.e. $\mathfrak{V}_m$ is the representation space [6, pp. 96–98] for the Kronecker m^{th} power representation of $\mathfrak{G}$. Let s be the permutation $1 \to 1', \cdots, m \to m'$. We define s as a linear operator on $\mathfrak{V}_m$ by the equation $x' = sx$ where $x'(i_1, \cdots, i_m) = x(i_{1'}, \cdots, i_{m'})$. Let $T_m(s)$ be the matrix which describes this mapping. It is evident that $\mathfrak{T}_m : s \to T_m(s)$ is a permutation representation of degree n^m of $\mathfrak{S}_m$.

We call a $\mathfrak{k}$-matrix of degree n^m *bisymmetric* if it commutes with every $T_m(s)$. Since the set $\mathfrak{B}_m$ of all bisymmetric $\mathfrak{k}$-matrices is the commutator algebra of a permutation representation, its order will be the same as the order of the set $\mathfrak{B}_m^0$ of all bisymmetric matrices in a field of characteristic 0. This latter order [6, p. 130] is known to be $\binom{N+m-1}{m}$.

It is trivial to verify that $\Pi_m(A)$ is bisymmetric; i.e. $\mathfrak{A}_m \subseteq \mathfrak{B}_m$. Now apply Lemma IV and we see that

THEOREM III. *If $\mathfrak{k}$ has at least m elements, then the enveloping algebra of the Kronecker m^{th} power representation of the full linear group of degree n is the set of all bisymmetric $\mathfrak{k}$-matrices of degree n^m.*

Still paralleling the non-modular theory, our next step is to determine the indecomposable constituents of $\mathfrak{T}_m$. Then we obtain the decomposed form of $\mathfrak{A}_m$, by starting with the commutator algebra of the decomposed form of $\mathfrak{T}_m$. When this is all accomplished we shall know the structure of the decomposed form of Π_m; i.e. we shall know the degrees of the irreducible constituents; the nature and multiplicities of the indecomposable constituents. But we shall still have no direct construction for the representations themselves, or much information about the characters of the representations. This is a general difficulty encountered when a representation of a group is studied by determining its enveloping algebra as a commutator algebra. The root of this difficulty is the lack of criteria for determining which elements of the enveloping algebra actually correspond to group elements. Attempts to remedy these deficiencies for the present theory are postponed to later papers. We also postpone any discussion of the case in which $\mathfrak{k}$ has less than m elements.

5. The representations $\mathfrak{T}_m$

We may regard $\mathfrak{T}_m$ as a permutation group whose elements $T_m(s)$ are written on the "letters" $x(i_1, \cdots, i_m)$. Two letters $x(i_1, \cdots, i_m)$ and $x(j_1, \cdots, j_m)$ belong to the same system of transitivity of $\mathfrak{T}_m$ if and only if the integers $j_1, \cdots, j_m$ are just the integers $i_1, \cdots, i_m$ in some arrangement. We now arrange the basis vectors of $\mathfrak{V}_m$ so that the letters of the several systems of transitivity are brought together. In the language of representation theory,

this exhibits the representation $\mathfrak{T}_m$ as the direct sum [6, pp. 19, 20] of its *transitive constituents*.

Let $(\lambda) = (\lambda_1, \cdots, \lambda_k)$ denote the partition $m = \lambda_1 + \cdots + \lambda_k$, $\lambda_1 \geqq \cdots \geqq \lambda_k > 0$, and let $x(i_1, \cdots, i_m)$ have the first λ_1 indices all $1, \cdots$, the last λ_k indices all k. The permutations s such that $sx(i_1, \cdots, i_m) = x(i_1, \cdots, i_m)$ constitute a subgroup of $\mathfrak{S}_m$ isomorphic to $\mathfrak{S}(\lambda) = \mathfrak{S}_{\lambda_1} \times \cdots \times \mathfrak{S}_{\lambda_k}$ (here $\times$ denotes group-theoretic direct product). Let $S(\lambda)$ denote the sum of the elements of $\mathfrak{S}(\lambda)$, and suppose that $s_1 = 1, s_2, \cdots$ are elements, one from each coset of $\mathfrak{S}(\lambda)$ in $\mathfrak{S}_m$. Then the elements $s_i S(\lambda)$ constitute a $\mathfrak{k}$-basis for the left ideal $\mathfrak{L}(\lambda)$ of $\mathfrak{R}_m$ (the $\mathfrak{k}$ group ring of $\mathfrak{S}_m$) generated by $S(\lambda)$. Left multiplication of $\mathfrak{L}(\lambda)$ by a permutation s merely permutes the basis elements $s_i S(\lambda)$, and so $\mathfrak{L}(\lambda)$ is representation space for a (transitive) permutation representation [3, p. 110] $\mathfrak{T}_{(\lambda)} : s \rightarrow T_{(\lambda)}(s)$ of $\mathfrak{S}_m$.

It is obvious that, as left $\mathfrak{R}_m$-space $\mathfrak{L}(\lambda)$ is isomorphic to the $\mathfrak{k}$-space made up of tensors whose only non-zero coordinates are $x(i_1, \cdots, i_m)$ and its conjugates $sx(i_1, \cdots, i_m)$. Hence $\mathfrak{T}_{(\lambda)}$ is equivalent to a constituent of $\mathfrak{T}_m$; and, conversely, it is clear that every transitive constituent of $\mathfrak{T}_m$ is equivalent to one of the $\mathfrak{T}_{(\lambda)}$. So to know the decomposition of $\mathfrak{T}_m$ we need only know the decomposition of each $\mathfrak{T}_{(\lambda)}$; or equivalently, to write each $\mathfrak{L}(\lambda)$ as a direct sum of indecomposable left ideals of $\mathfrak{R}_m$.

We have $S(\lambda)^2 = n(\lambda)S(\lambda)$ where $n(\lambda) = \lambda_1! \cdots \lambda_k!$. Hence if λ_1 (and therefore every λ_i) is less than p, the ideal $\mathfrak{L}(\lambda)$ has the idempotent generator $e(\lambda) = S(\lambda)/n(\lambda)$; and so[2] $\mathfrak{L}(\lambda) = \mathfrak{R}_m S(\lambda)$ can be written as the direct sum of indecomposable left ideals which are direct summands of $\mathfrak{R}_m$ (i.e. which themselves have idempotent generators). Stating this in the language of representation theory we have

THEOREM IV. *If $\lambda_1 < p$, $\mathfrak{T}_{(\lambda)}$ is a direct sum of indecomposable constituents of the regular representation of $\mathfrak{S}_m$.*

If $m = p$, Theorem IV covers all but one partition, the exception being $\lambda_1 = p$. But $\mathfrak{T}_{(p)}$ is the identity representation, and so we have established the following theorem for $m = p$:

THEOREM V. *For $m < 2p$ an indecomposable constituent of $\mathfrak{T}_m$ is either an indecomposable constituent of the regular representation of $\mathfrak{S}_m$, or one of the irreducible representations*[3] $\mathfrak{F}_1(\mu)$ *of $\mathfrak{S}_m$.*

PROOF. There is nothing to prove for $m < p$. We proceed by an induction on m based upon the already verified case $m = p$. We suppose $p < m < 2p$ and that the theorem is already verified for $m - 1$. Considered only for elements of $\mathfrak{S}_{m-1}$, $\mathfrak{T}_m$ is just $\mathfrak{T}_{m-1}$ repeated n times. Hence, any indecomposable constituent of $\mathfrak{T}_m$ must, when considered only for elements of $\mathfrak{S}_{m-1}$, split into indecomposable constituents of $\mathfrak{T}_{m-1}$.

Reference to Theorem I shows that Theorem V is false only if (I) $\mathfrak{T}_m$ has

[2] See Theorem I [1], p. 9.

[3] The notation $\mathfrak{F}_1(\mu)$ is explained in [5], §6.

$\mathfrak{F}_j(\mu)$, $j > 1$, as an indecomposable direct constituent, or (II) $\mathfrak{T}_m$ has an indecomposable direct constituent of Loewy length 2. We now apply Corollary I to show that either I or II contradicts our induction hypothesis.

The application to I is immediate. For let (μ) be a partition of $l = m - p$. Then since $m > p$, there will be some i for which $\mu_i > \mu_{i+1}$; and so $\mathfrak{F}_j(\mu)$ in $\mathfrak{T}_m$ would require $\mathfrak{F}_j(\mu \mid i)$ in $\mathfrak{T}_{m-1}$, contrary to our induction hypothesis.

Let $\mathfrak{B}$ be an indecomposable representation of $\mathfrak{S}_m$ of Loewy length 2, whose irreducible constituents are $\mathfrak{F}_j(\mu)$, $j = j_0, \cdots, j_0 + r$, $r \geqq 1$. Let $\mathfrak{B}^*$ denote $\mathfrak{B}$ considered only for elements of $\mathfrak{S}_{m-1}$. If $\mathfrak{B}$ is an indecomposable direct constituent of $\mathfrak{T}_m$, then $\mathfrak{B}^*$ is a direct constituent of $\mathfrak{T}_{m-1}$.

By Corollary I, the irreducible constituents of $\mathfrak{B}^*$ will be either of highest kind or of the form $\mathfrak{F}_j(\mu \mid i)$ for j in the range $j_0, \cdots, j_0 + r$. Since no $\mathfrak{F}_j(\mu)$ is repeated in $\mathfrak{B}$, it follows from Corollary I that no $\mathfrak{F}_j(\mu \mid i)$ can be repeated in any indecomposable consistuent of $\mathfrak{B}^*$; and so $\mathfrak{B}^*$ can contain no indecomposable constituent of Loewy length 3. Since $r \geqq 1$ and $m > p$, $\mathfrak{B}^*$ must contain some $\mathfrak{F}_j(\mu \mid i)$ for $j > 1$. This $\mathfrak{F}_j(\mu \mid i)$ must lie in an indecomposable direct constituent of $\mathfrak{B}^*$ of Loewy length 1 or 2. But then $\mathfrak{B}^*$ cannot be a direct constituent of $\mathfrak{T}_{m-1}$, because of our induction hypothesis; hence $\mathfrak{B}$ cannot be a direct constituent of $\mathfrak{T}_m$; i.e. II is impossible. This completes the proof of Theorem V.

6. The commutator algebra of $\mathfrak{T}_m$

Instead of studying $\mathfrak{T}_m$ itself, we investigate the more general case of any representation which is the sum of any number of the representations $\mathfrak{T}_{(\lambda)}$, repetitions permitted. Let $\mathfrak{B}$ denote the decomposed form of any such representation. We group the constituents of $\mathfrak{B}$ in such a way that

$$\mathfrak{B} = \left\| \begin{matrix} \mathfrak{B}_1 & & \\ & \mathfrak{B}_2 & \\ & & \ddots \end{matrix} \right\| \tag{28}$$

where $\mathfrak{B}_r$ consists of all the constituents of $\mathfrak{B}$ that belong to the block $\mathfrak{B}_r$ of $\mathfrak{R}_m$. The blocks $\mathfrak{B}_r$ correspond to two-sided ideals that are minimal direct summands of $\mathfrak{R}_m$. Hence the commutator algebra $\mathfrak{W}$ of $\mathfrak{B}$ is the direct sum of the commutator algebras $\mathfrak{W}_r$ of the $\mathfrak{B}_r$ i.e.

$$\mathfrak{W} = \left\| \begin{matrix} \mathfrak{W}_1 & & \\ & \mathfrak{W}_2 & \\ & & \ddots \end{matrix} \right\| \tag{29}$$

If $\mathfrak{B}_r$ belongs to a block of lowest kind, it will just be an $\mathfrak{F}(\lambda)$ repeated, say, δ times. Then,[4] since $\mathfrak{F}(\lambda)$ is a total matrix algebra, $\mathfrak{W}_r$ is equivalent to the

[4] Cf. [6], p. 92.

total $\mathfrak{k}$-matrix algebra of degree δ repeated $f(\lambda)$ times, where $f(\lambda)$ is the degree of $\mathfrak{F}(\lambda)$.

If $\mathfrak{B}_\tau$ belongs to a block $\mathfrak{B}_\tau = \mathfrak{B}(\mu)$ of highest kind, the situation is somewhat more complicated. For simplicity in notation we drop all arguments (μ) from the letters denoting representations and matrices. According to Theorem V, $\mathfrak{B}_\tau$ has the form:

$$\mathfrak{B}_\tau = \left\| \begin{matrix} E_{\gamma_0} \times \mathfrak{F}_1 & & & \\ & E_{\gamma_1} \times \mathfrak{U}_1 & & \\ & & \ddots & \\ & & & E_{\gamma_{p-1}} \times \mathfrak{U}_{p-1} \end{matrix} \right\| \tag{30}$$

where E_ν is the unit matrix of degree ν and $\times$ denotes Kronecker product. To obtain uniformity in notation we set $\mathfrak{F}_1 = \mathfrak{U}_0$. Suppose that $W_{ij}U_j(s) = U_i(s)W_{ij}$, for all s in $\mathfrak{S}_m$, and denote by A_{ij} any $\mathfrak{k}$-matrix of γ_i rows and γ_j columns, $i, j = 0, \cdots, p-1$. Then

$$W_\tau = \left\| \begin{matrix} A_{00} \times W_{00} \cdots A_{0p-1} \times W_{0p-1} \\ \cdots \\ A_{p-10} \times W_{p-10} \cdots A_{p-1p-1} \times W_{p-1p-1} \end{matrix} \right\| \tag{31}$$

is an element of $\mathfrak{W}_\tau$; and, conversely, any element of $\mathfrak{W}_\tau$ can be written as a linear $\mathfrak{k}$-combination of elements of the form (31).

To determine the number h_{ij} of linearly independent W_{ij} we use the Cartan matrix for the block $\mathfrak{B}$ and apply the general theory of intertwining matrices.[5] The result is that $h_{ij} = 0$ (and therefore $W_{ij} = 0$) unless j is $i - 1$, i, or $i + 1$. $h_{00} = 1$; $h_{ii} = 2$, $i = 1, \cdots, p-1$; $h_{i,i+1} = h_{i+1,i} = 1$, $i = 0, \cdots, p-1$. Since the A_{ij}'s are arbitrary this gives

$$\begin{aligned} &\gamma_0^2 + 2\gamma_0\gamma_1 + 2\gamma_1^2 + 2\gamma_1\gamma_2 + \cdots + 2\gamma_{p-1}^2 \\ &\qquad = (\gamma_0 + \gamma_1)^2 + (\gamma_1 + \gamma_2)^2 + \cdots + \gamma_{p-1}^2 \end{aligned} \tag{32}$$

for the order of $\mathfrak{W}_\tau$.

To determine the structure of $\mathfrak{W}_\tau$ we must know the form of all the W_{ij}. We subdivide W_{ij} so that the rows of its submatrices are the same as the rows occupied by the irreducible constituents of $\mathfrak{U}_i$ and the columns of the submatrices are the same as the rows occupied by the irreducible constituents of $\mathfrak{U}_j$. See formula (11) for the form of the $\mathfrak{U}_i$. We omit the details of the

[5] See, for instance, Theorem 4 [4], p. 648.

computation. f_i denotes the degree of $\mathfrak{F}_i$, 0 stands for the zero matrix of proper size.

$$W_{00} = \| aE_{f_1} \|, \qquad W_{01} = \| aE_{f_1} \quad 0 \quad 0 \|,$$

$$W_{10} = \begin{Vmatrix} 0 \\ 0 \\ aE_{f_1} \end{Vmatrix}, \qquad W_{ii} = \begin{Vmatrix} aE_{f_1} & 0 & 0 & 0 \\ 0 & aE_{f_{i-1}} & 0 & 0 \\ 0 & 0 & aE_{f_{i+1}} & 0 \\ bE_{f_i} & 0 & 0 & aE_{f_i} \end{Vmatrix}, \tag{33}$$

$$W_{ii+1} = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ aE_{f_{i+1}} & 0 & 0 & 0 \\ 0 & aE_{f_i} & 0 & 0 \end{Vmatrix}, \qquad W_{i+1,i} = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ aE_{f_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & aE_{f_{i+1}} & 0 \end{Vmatrix}$$

The a's appearing in different matrices W_{ij} are totally unrelated; for $i = 1$ and $i = p - 1$ the rows and columns of the W_{ij} that correspond formally to $\mathfrak{F}_0$ and $\mathfrak{F}_p$ are to be deleted.

After a suitable shifting of rows and columns, $\mathfrak{W}_r$ takes the form

$$\begin{Vmatrix} E_{f_1} \times \mathfrak{W}^1 & & \\ & \ddots & \\ & & E_{f_{p-1}} \times \mathfrak{W}^{p-1} \end{Vmatrix} \tag{34}$$

where $\mathfrak{W}^i$ is the indecomposable matrix algebra given by

$$\mathfrak{W}^i = \begin{Vmatrix} \mathfrak{F}^i & & & \\ \mathfrak{F}_i^{i-1} & \mathfrak{F}^{i-1} & & \\ \mathfrak{F}_i^{i+1} & 0 & \mathfrak{F}^{i+1} & \\ \mathfrak{F}_i^i & \mathfrak{F}_{i-1}^i & \mathfrak{F}_{i+1}^i & \mathfrak{F}^i \end{Vmatrix} \qquad i = 1, \cdots, p-1. \tag{35}$$

(For $i = p - 1$ the rows and columns of $\mathfrak{F}^p$ are deleted.) Each $\mathfrak{F}^i$ is a total $\mathfrak{k}$-matrix algebra of degree γ_i and the $\mathfrak{F}_j^i$ are total $\mathfrak{k}$-matrix sets of degrees indicated by their positions in $\mathfrak{W}^i$.

It is of some interest to observe that if (I): $\gamma_i \neq 0$ implies $\gamma_\nu \neq 0$ for all $\nu < i$, holds for $\mathfrak{B}_r$ then the commutator algebra of $\mathfrak{W}_r$ is just the enveloping algebra of $\mathfrak{B}_r$.

7. Multiplicities of the constituents of $\mathfrak{T}_m$

We continue to regard $\mathfrak{T}_{(\lambda)}$, $\mathfrak{T}_m$, as permutation representations of $\mathfrak{S}_m$ with the given $\mathfrak{k}$ of characteristic p as underlying field. We shall denote by $\mathfrak{T}^0_{(\lambda)}$, $\mathfrak{T}^0_m$ the same permutation representations of $\mathfrak{S}_m$, but now regarded as made

up of matrices in a field $\mathfrak{K}$ of characteristic 0. The structure of the representations $\mathfrak{T}^0_{(\lambda)}$, $\mathfrak{T}^0_m$ is well known.[6] Denote by $\delta(\lambda)$ the multiplicity of $\mathfrak{A}(\lambda)$ in T^0_m; by $\delta(\lambda \between \lambda')$ the multiplicity of $\mathfrak{A}(\lambda)$ in $\mathfrak{T}^0_{(\lambda')}$.

If the diagram of (λ) has no hook of length p, i.e. if $\mathfrak{F}(\lambda)$ is of highest kind, then $\mathfrak{F}(\lambda)$ occurs $\delta(\lambda)$ times in $\mathfrak{T}_m$ and $\delta(\lambda \between \lambda')$ times in $\mathfrak{T}_{(\lambda')}$. If the diagram of (λ) has a hook of length p, say $(\lambda) = \lambda^i(\mu)$ then we set $\delta(\lambda) = \delta_i(\mu)$ and $\delta(\lambda \between \lambda') = \delta_i(\mu \between \lambda')$. We let $\gamma_0(\mu), \gamma_1(\mu), \cdots, \gamma_{p-1}(\mu)$ denote the multiplicities in $\mathfrak{T}_m$ of $\mathfrak{F}_1(\mu), \mathfrak{U}_1(\mu), \cdots, \mathfrak{U}_{p-1}(\mu)$ respectively; and let $\gamma_0(\mu \between \lambda'), \cdots, \gamma_{p-1}(\mu \between \lambda')$ denote the corresponding multiplicies in $\mathfrak{T}_{(\lambda')}$. In formulas where they appear formally we set $\gamma_p(\mu) = \gamma_p(\mu \between \lambda) = 0$.

We get relations between the δ's and the γ's by counting the multiplicity of the modular irreducible constituents in two ways. Considering $\mathfrak{T}_m$ as the modular representation induced by $\mathfrak{T}^0_m$, it follows from the form of the decomposition matrix[7] $D(\mu)$ for the block $\mathfrak{B}(\mu)$, that $\mathfrak{F}_i(\mu)$ occurs $\delta_i(\mu) + \delta_{i+1}(\mu)$ times in $\mathfrak{T}_m$. On the other hand, using the Cartan matrix $C(\mu)$ for the block $\mathfrak{B}(\mu)$ we see that $\mathfrak{F}_i(\mu)$ occurs $\gamma_{i-1}(\mu) + 2\gamma_i(\mu) + \gamma_{i+1}(\mu)$ times in $\mathfrak{T}_m$. Hence

$$(36) \qquad \delta_i(\mu) + \delta_{i+1}(\mu) = \gamma_{i-1}(\mu) + 2\gamma_i(\mu) + \gamma_{i+1}(\mu), \qquad i = 1, \cdots, p-1.$$

The same reasoning applies to $\mathfrak{T}_{(\lambda)}$ giving

$$(37) \qquad \begin{aligned} \delta_i(\mu \between \lambda) + \delta_{i+1}(\mu \between \lambda) = \gamma_{i-1}(\mu \between \lambda) &+ 2\gamma_i(\mu \between \lambda) \\ &+ \gamma_{i+1}(\mu \between \lambda), \qquad i = 1, \cdots, p-1. \end{aligned}$$

We can now state the main theorem on multiplicities:

THEOREM VI. *If the diagram of* (λ) *has no hook of length* p *then* $\mathfrak{F}(\lambda)$ *occurs exactly as often in* $\mathfrak{T}_m$ *as* $\mathfrak{A}(\lambda)$ *occurs in* $\mathfrak{T}^0_m$. *For a block* $\mathfrak{B}(\mu)$ *of lowest kind we have*:

$$(38) \qquad \begin{aligned} \gamma_{p-1}(\mu) &= \delta_p(\mu) \\ \gamma_{p-2}(\mu) &= \delta_{p-1}(\mu) - \delta_p(\mu) \\ &\cdots\cdots\cdots\cdots \\ \gamma_{p-i}(\mu) &= \delta_{p-i+1}(\mu) - \delta_{p-i}(\mu) + \cdots + (-1)^{i+1}\delta_p(\mu). \end{aligned}$$

Only the second part requires additional proof. If we add to equations (36) the single equation

$$(39) \qquad \delta_1(\mu) = \gamma_0(\mu) + \gamma_1(\mu)$$

then a simple computation shows that (38) is the only solution of the augmented system. Hence to prove our theorem it is sufficient to establish (39). We do this by showing that

$$(40) \qquad \delta_1(\mu \between \lambda) = \gamma_0(\mu \between \lambda) + \gamma_1(\mu \between \lambda)$$

[6] [3], Chapter IV, especially pp. 110, 128, 129; [6], Chapter VII, §§5, 6, 7.

[7] [5], formulas (3) and (4).

for every partition (λ) of m, and then using the fact that $\mathfrak{T}_m$ is a direct sum of constituents $\mathfrak{T}_{(\lambda)}$. Of course (40) in the presence of (37) leads to the following analogue of (38):

$$\gamma_{p-i}(\mu)(\lambda) = \delta_{p-i+1}(\mu)(\lambda) - \delta_{p-i}(\mu)(\lambda) + \cdots + (-1)^{i+1}\delta_p(\mu)(\lambda), \quad i = 1, \cdots, p. \tag{41}$$

We arrange the partitions (λ) of m and the partitions (μ) of $l = m - p$ in dictionary order (i.e. (λ) precedes (λ') if the first non-vanishing difference $\lambda_i - \lambda_i'$ is positive). The verification of (40) is an induction argument, along the following lines: We suppose that (40) has been established for all $\mathfrak{T}_{(\lambda)}$ provided (μ) precedes a given (μ^0). Then we exhibit one (λ^0) (actually $\lambda^1(\mu^0)$) such that $\mathfrak{F}_1(\mu^0)$ is the only constituent of $\mathfrak{B}(\mu^0)$ which appears in $\mathfrak{T}_{(\lambda^0)}$. Let $N(*)$ denote the order of the commutator algebra of any representation, $*$, of $\mathfrak{S}_m$. Then by Remark I we have

$$N(\mathfrak{B}) = N(\mathfrak{B}^0) \tag{42}$$

where $\mathfrak{B}$ stands for any sum of $\mathfrak{T}_{(\lambda)}$ and $\mathfrak{B}^0$ is the sum of the corresponding $\mathfrak{T}^0_{(\lambda)}$. We establish (40) by solving for $\delta_1(\mu^0)(\lambda)$ in the equations obtained from (42) by setting $\mathfrak{B} = \mathfrak{T}_{(\lambda)}, \mathfrak{T}_{(\lambda^0)}, \mathfrak{T}_{(\lambda)} + \mathfrak{T}_{(\lambda^0)}$ in turn.

An important step in this process is the proof that a suitable (λ^0) exists. To accomplish this we analyze the character[8] of $\mathfrak{T}^0_{(\lambda')}$, where (λ') is any partition of m, and obtain the following

LEMMA V. *$\mathfrak{A}(\lambda')$ occurs exactly once as a constituent of $\mathfrak{T}^0_{(\lambda')}$, and $\mathfrak{A}(\lambda)$ cannot be a constituent of $\mathfrak{T}^0_{(\lambda')}$ if (λ') precedes (λ); i.e.*

$$\delta(\lambda')(\lambda') = 1, \qquad \delta(\lambda)(\lambda') = 0 \quad \textit{unless } (\lambda) \textit{ precedes } (\lambda'). \tag{43}$$

We observe that if (μ^0) precedes (μ) then $\lambda^1(\mu^0)$ precedes $\lambda^1(\mu)$, and for any (μ), $\lambda^1(\mu)$ precedes $\lambda^i(\mu)$ if $i > 1$. Then from (43) and (37) we get (since the γ's and δ's are non-negative integers),

LEMMA VI. *Let $(\lambda^0) = \lambda^1(\mu^0) = (\mu_1^0 + p, \mu_2^0, \cdots)$. Then* (i) *if (μ^0) precedes (μ) $\mathfrak{T}_{(\lambda^0)}$ contains no constituents from the block $\mathfrak{B}(\mu)$, and* (ii) *$\mathfrak{F}_1(\mu^0)$ is the only constituent of $\mathfrak{T}_{(\lambda^0)}$ belonging to the block $\mathfrak{B}(\mu^0)$ and it occurs with multiplicity one (i.e. $\gamma_0(\mu^0)(\lambda^0) = 1$).*

In §6 we saw that the commutator algebra of a representation $\mathfrak{B} = \sum_\rho \mathfrak{T}_{(\lambda^\rho)}$ can be computed block at a time. Let $N(\mathfrak{B}, (\mu))$ denote the contribution to $N(\mathfrak{B})$ of a block of lowest kind and $N(\mathfrak{B}, (\lambda))$ the same for a block of highest kind. Then (see formula (32)) we have

$$\begin{aligned} N(\mathfrak{B}, (\mu)) &= \textstyle\sum_{i=1}^{p} \left(\sum_\rho \gamma_{i-1}(\mu)(\lambda^\rho) + \gamma_i(\mu)(\lambda^\rho)\right)^2; \\ N(\mathfrak{B}, (\lambda)) &= \left(\textstyle\sum_\rho \delta(\lambda)(\lambda^\rho)\right)^2 \end{aligned} \tag{44}$$

and for the total order

$$N(\mathfrak{B}) = \sum N(\mathfrak{B}, (\mu)) + \sum N(\mathfrak{B}, (\lambda)), \tag{45}$$

---

[8] [2], pp. 71, 94; [3], p. 110; [6], p. 205.

where the first sum is over all partitions (μ) of l and the second sum is over all partitions (λ) of m whose diagrams have no hook of length p.

For the corresponding non-modular representation $\mathfrak{B}^0$ we have analogously

$$N(\mathfrak{B}^0, (\lambda)) = \left(\textstyle\sum_\rho \delta(\lambda \between \lambda^\rho)\right)^2,$$

and $N(\mathfrak{B}^0) = \sum N(\mathfrak{B}^0, (\lambda))$, the sum extending over *all* partitions (λ) of m. For comparison with $N(\mathfrak{B})$ it is convenient to group together those $N(\mathfrak{B}^0, (\lambda))$ for which the representations $\mathfrak{F}(\lambda)$ belong to the same blocks. Thus we obtain

$$N(\mathfrak{B}^0) = \sum N(\mathfrak{B}^0, (\mu)) + \sum N(\mathfrak{B}^0, (\lambda)) \tag{46}$$

where the summation ranges are the same as in (45) and

$$N(\mathfrak{B}^0, (\mu)) = \textstyle\sum_{i=1}^{p} N(\mathfrak{B}^0, \lambda^i(\mu)) = \sum_{i=1}^{p} \left(\sum_\rho \delta_i(\mu \between \lambda^\rho)\right)^2. \tag{47}$$

By Remark I the difference $N(\mathfrak{B}^0) - N(\mathfrak{B})$ is zero. Observe that the $N(\mathfrak{B}^0, (\lambda))$ and $N(\mathfrak{B}, (\lambda))$ from (45) and (46) cancel. Furthermore, by our induction hypothesis, (40) holds for any (μ) which precedes (μ^0); this in turn implies that $N(\mathfrak{B}^0, (\mu)) = N(\mathfrak{B}, (\mu))$ for any (μ) which precedes (μ^0). Hence we have

$$0 = \sum N(\mathfrak{B}^0, (\mu)) - N(\mathfrak{B}, (\mu)) \tag{48}$$

where the summation extends over all (μ) (including (μ^0)) which do not precede (μ^0).

Now let (λ) be any partition of m. By (43), (47), $N(\mathfrak{T}^0_{(\lambda)} + \mathfrak{T}^0_{(\lambda^0)}, (\mu)) = N(\mathfrak{T}^0_{(\lambda)}, (\mu))$ if (μ) follows (μ^0). By Lemma VI and (44), $N(\mathfrak{T}_{(\lambda)} + \mathfrak{T}_{(\lambda^0)}, (\mu)) = N(\mathfrak{T}_{(\lambda)}, (\mu))$ if (μ) follows (μ^0). Hence if we subtract (48) for $\mathfrak{B} = \mathfrak{T}_{(\lambda)}$ from (48) for $\mathfrak{B} = \mathfrak{T}_{(\lambda)} + \mathfrak{T}_{(\lambda^0)}$ the only terms which do not cancel are those involving (μ^0). This leads us to

$$\begin{aligned} N(\mathfrak{T}^0_{(\lambda)} + \mathfrak{T}^0_{(\lambda^0)}, (\mu^0)) - N(\mathfrak{T}^0_{(\lambda)}(\mu^0)) \\ = N(\mathfrak{T}_{(\lambda)} + \mathfrak{T}_{(\lambda^0)}, (\mu^0)) - N(\mathfrak{T}_{(\lambda)}, (\mu^0)). \end{aligned} \tag{49}$$

Now substituting in this from (44) and (47) and referring to (43) and Lemma VI for the values of $\delta_i(\mu^0 \between \lambda^0)$, $\gamma_i(\mu^0 \between \lambda^0)$ we have

$$\begin{aligned} &[(\delta_1(\mu^0 \between \lambda) + 1)^2 + \delta_2(\mu^0 \between \lambda)^2 + \cdots + \delta_p(\mu^0 \between \lambda)^2] - [\delta_1(\mu^0 \between \lambda)^2 + \cdots + \delta_p(\mu^0 \between \lambda)^2] \\ &= [(\gamma_0(\mu^0 \between \lambda) + \gamma_1(\mu^0 \between \lambda) + 1)^2 + (\gamma_1(\mu^0 \between \lambda) + \gamma_2(\mu^0 \between \lambda))^2 + \cdots + \gamma_{p-1}(\mu^0 \between \lambda)^2] \\ &\quad - [(\gamma_0(\mu^0 \between \lambda) + \gamma_1(\mu^0 \between \lambda))^2 + \cdots + \gamma_{p-1}(\mu^0 \between \lambda)^2] \end{aligned}$$

or

$$2\delta_1(\mu^0 \between \lambda) + 1 = 2(\gamma_0(\mu^0 \between \lambda) + \gamma_1(\mu^0 \between \lambda)) + 1$$

which is the same as (40). Observe that the argument above applies to the first partition, $\mu_1 = l$ of l; hence our induction is complete and Theorem VI is fully established.

In words (49) says that the change in order of the commutator algebra, due to addition of one particular permutation representation to another particular permutation representation, is the same, block by block, characteristic 0 as characteristic p. Remark I states merely that the total change is the same. If we could strengthen Remark I to a block by block form, the proof of the above theorem could be much shortened, as then one could omit everything between Lemma VI and formula (49); and of course any such improvement of Remark I would be of interest in the general modular theory entirely aside from its application here.

8. The Kronecker m^{th} power of the full linear group

In order to describe the structure of the enveloping algebra $\mathfrak{A}_m$ of Π_m we have only to put together the results of the preceding sections and introduce a suitable notation for the constituents.

Let (λ) be a partition of m whose diagram has no hook of length p. Then we denote by $\mathfrak{G}(\lambda)$ the total $\mathfrak{k}$-matrix algebra of degree $\delta(\lambda)$. Let (μ) be a partition of $l = m - p$. Then we denote by $\mathfrak{G}_i(\mu)$ the total $\mathfrak{k}$-matrix algebra of degree[9] $\gamma_i(\mu)$, and by $\mathfrak{G}_j^i(\mu)$, for $j = i - 1, i, i + 1$, the set of all $\gamma_i(\mu)$ by $\gamma_j(\mu)$ $\mathfrak{k}$-matrices; all this for $i = 0, \cdots, p - 1$, with the usual conventions for $i = 0, i = p - 1$ (i.e., $j < 0, j > p - 1$ are excluded). Finally, we set

$$\tag{50} \mathfrak{U}^i(\mu) = \left\| \begin{matrix} \mathfrak{G}_i(\mu) & & & \\ \mathfrak{G}_i^{i-1}(\mu) & \mathfrak{G}_{i-1}(\mu) & & \\ \mathfrak{G}_i^{i+1}(\mu) & 0 & \mathfrak{G}_{i+1}(\mu) & \\ \mathfrak{G}_i^i(\mu) & \mathfrak{G}_{i-1}^i(\mu) & \mathfrak{G}_{i+1}^i(\mu) & \mathfrak{G}_i(\mu) \end{matrix} \right\| \qquad i = 1, \cdots, p - 1.$$

We can now state the main theorem on the structure of $\mathfrak{A}_m$.

THEOREM VII. *For $m < 2p$ and $\mathfrak{k}$ any field (of characteristic p) containing at least m elements, the enveloping algebra $\mathfrak{A}_m$, of the Kronecker m^{th} power representation Π_m, of the full linear group $\mathfrak{G}$, of n-rowed non-singular $\mathfrak{k}$-matrices, has the following indecomposable (direct) constituents*: (i) *If (λ) is any partition of m whose diagram has no hook of length p, then $\mathfrak{G}(\lambda)$ appears $f(\lambda)$ times as an indecomposable constituents of $\mathfrak{A}_m$; where $f(\lambda)$ is the degree of the ordinary irreducible representation of $\mathfrak{S}_m$ defined by (λ).* (ii) *If (μ) is any partition of $l = m - p$, then $\mathfrak{U}^i(\mu)$ appears $f_i(\mu)$ times as an indecomposable constituent of $\mathfrak{A}_m$, $i = 1, \cdots, p - 1$, where $f_i(\mu)$ is the degree of the irreducible modular representation $\mathfrak{F}_i(\mu)$ of $\mathfrak{S}_m$.*

The theorem follows from Theorem III, the formulas of §6, and Theorem VI. The degrees $f(\lambda)$ are well known [6, p. 213], and for the [10] $f_i(\mu)$ we have

$$\tag{51} f_i(\mu) = f(\lambda^i(\mu)) - f(\lambda^{i-1}(\mu)) + \cdots + (-1)^{i+1} f(\lambda^1(\mu))$$

---

[9] See formula (38) for the value of $\gamma_i(\mu)$.

[10] Cf. [5], formula (20).

where[11] $\lambda^j(\mu)$ is the partition of m whose diagram T has a hook H, of length p and height (vertical length) j, such that $T - H$ is the diagram of (μ).

Any indecomposable constituent of $\mathfrak{A}_m$ affords an indecomposable representation of $\mathfrak{G}$. We shall consider the symbols $\mathfrak{G}(\lambda)$, $\mathfrak{U}^i(\mu)$, $\mathfrak{G}_i(\mu)$ in two ways: first, as they are defined above; and second, as denoting representations of $\mathfrak{G}$; for instance $\mathfrak{G}(\lambda)$ is the representation $A \rightarrow G(\lambda\between A)$, where $G(\lambda\between A)$ is the matrix in $\mathfrak{G}(\lambda)$ assigned to $\Pi_m(A)$ considered as an element of $\mathfrak{A}_m$. We define the matrices $U^i(\mu\between A)$, $G_i^j(\mu\between A)$ analogously. Then *the* $\mathfrak{G}(\lambda)$, $\mathfrak{G}_i(\mu)$ *are all the irreducible representations of* $\mathfrak{G}$ *which are induced in the space of tensors of rank* m.

INSTITUTE FOR ADVANCED STUDY

BIBLIOGRAPHY

[1] M. DEURING, *Algebren*, Berlin, 1935.
[2] D. E. LITTLEWOOD, *The Theory of Group Characters*, Oxford, 1940.
[3] F. MURNAGHAN, *The Theory of Group Representations*, Baltimore, 1938.
[4] C. J. NESBITT, *On the Regular Representations of Algebras*, these Annals, vol. 39 (1938), pp. 634–658.
[5] R. M. THRALL AND C. J. NESBITT, *On the Modular Representations of the Symmetric Group*, these Annals, vol. 43 (1942), pp. 656–670.
[6] H. WEYL, *The Classical Groups*, Princeton, 1939.

[11] See [5], §6, for a discussion of hooks and for references to Nakayama's treatment.

Annals of Mathematics
Vol. 43, No. 4, October, 1942

NON-ASSOCIATIVE ALGEBRAS[1]

I. Fundamental Concepts and Isotopy

By A. A. Albert

(Received January 5, 1942)

1. Introduction

The study of non-associative algebras has already yielded much of interest and importance. Indeed those special theories[2] in which the associative law is replaced by a substitute have each been of an extent and of an interest almost comparable to that of the theory of associative algebras.

The results on non-associative algebras in which one does not assume a type of partial associativity[3] have almost[4] all been of a rather primitive kind and have been scattered through the literature. They have, in particular, not emphasized adequately the important fact that many of the properties of arbitrary linear algebras are equivalent to certain properties of related sets of linear transformations in which multiplication then does satisfy the associative law. The fact that there is a rather surprisingly large number of non-associative algebras of orders two and three has been noted[3] but it has not been recognized before that this is at least partly due to the undesirable narrowness of the concept of equivalence for algebras other than associative algebras with a unity quantity.

It is the purpose of that part of the study of non-associative algebras which we shall begin here to emphasize the facts noted above by providing an appropriate formulation of the fundamentals of the theory of arbitrary linear algebras. Thus we shall devote the first portion of our present discussion to the process of relating the elementary properties of an algebra $\mathfrak{A}$ to the corresponding properties of three attached linear spaces of linear transformations on $\mathfrak{A}$. We shall then introduce the concept of *isotopy* of algebras, an extension of the concept of equivalence which coincides with the latter concept in the case of associative algebras with a unity quantity. Our discussion will conclude with an extensive

---

[1] Presented to the Society September 5, 1941. Most of the results of this paper were announced also in lectures at Princeton and Harvard in March 1941.

[2] We refer here first to the theory of alternative algebras for which see M. Zorn, *Theorie der Alternative Ringe*, Hamburg Abh., vol. 8 (1930), pp. 123–47, *Alternativekörper und Quadratische Systeme*, loc. cit., vol. 9 (1933), pp. 395–402. A second such theory is that of Lie Algebras for which see N. Jacobson, *Simple Lie algebras over a field of characteristic zero*, Duke J., vol. 4 (1938), pp. 534–51. Finally Jordan algebras are described in P. Jordan, J. v. Neumann, and E. Wigner, *On an algebraic generalization of the quantum mechanical formalism*, these Annals, vol. 35 (1934), pp. 29–64. The articles quoted contain bibliographies of their subjects and it should be remarked here that the theory of Jordan algebras has been generalized by G. Kalisch in his Chicago doctoral dissertation.

[3] Cf. L. E. Dickson, *Linear algebras with associativity not assumed*, Duke J., vol. 1 (1935), pp. 113–25.

[4] For a paper not of this type see footnote 6.

consideration of the question as to what properties of an algebra are preserved when we pass to an isotope.

It is the author's hope that the study begun here may ultimately lead to a solution of the problem of determining all simple algebras with a unity quantity, at least in a sense like that in which we say that the corresponding problem for associative algebras has been solved. A fundamental part of the associative algebra theory consisted of the definition of the known types of algebras, that is, the cyclic algebras and the crossed products, and such definitions will also be required for the non-associative case. We shall provide at least a very extensive part of this requirement in Part II of the present study. There we shall define[5] a class of non-commutative simple algebras with a unity quantity containing all such algebras which have been considered thus far in the literature as well as a very rich variety of new types.

2. The multiplication spaces of an algebra

A *linear algebra* of order n over a field $\mathfrak{F}$ is, in particular, a *linear space* of order n over $\mathfrak{F}$. But all linear spaces of the same order are equivalent. Thus we may regard all algebras of the same order as having the same quantities but with different laws for forming products. In particular we may take our quantities to be vectors, that is, one by n matrices

$$a = (\alpha_1, \cdots, \alpha_n) \qquad (\alpha_i \text{ in } \mathfrak{F}). \tag{1}$$

An algebra $\mathfrak{A}$ now consists of the linear space $\mathfrak{L}$ of all the vectors (1) together with a set of n^3 quantities γ_{ijk} in $\mathfrak{F}$ such that the product

$$u = a \cdot x = (\alpha_1, \cdots, \alpha_n) \cdot (\xi_1, \cdots, \xi_n) = (\mu_1, \cdots, \mu_n) \tag{2}$$

of any two quantities of $\mathfrak{L}$ is defined in $\mathfrak{A}$ by

$$\mu_k = \sum_{i,j=1}^{n} \alpha_i \gamma_{ijk} \xi_j \qquad (k = 1, \cdots, n). \tag{3}$$

Define $\Gamma^{(j)}$ to be the n-rowed square matrix with γ_{ijk} in the i^{th} row and k^{th} column, and write

$$\Gamma_x = \Gamma^{(1)}\xi_1 + \cdots + \Gamma^{(n)}\xi_n, \tag{4}$$

so that $\Gamma^{(i)} = \Gamma_{e_i}$ where e_i is given by (1) with $\alpha_i = 1$ and all the other $\alpha_j = 0$. The customary row by column definition of a matrix product does not include the definition of ax and so $a \cdot x \neq ax$. However it is clear that

$$a \cdot x = a\Gamma_x, \tag{5}$$

where $a\Gamma_x$ is computed as usual. Matrix multiplication is associative and so

$$(a \cdot x) \cdot y = (a\Gamma_x)\Gamma_y = a(\Gamma_x\Gamma_y). \tag{6}$$

[5] This definition has already been presented by the author in a lecture at the University of Cincinnati, November 15, 1941. See also the author's paper on *Quadratic forms permitting composition*, these Annals, this volume.

But

$$a \cdot (x \cdot y) = a\Gamma_u , \qquad u = x \cdot y = x\Gamma_y , \tag{7}$$

and $\mathfrak{A}$ is associative if and only if $\Gamma_x\Gamma_y = \Gamma_u$ for every x and y. We shall obtain another criterion later.

The linear transformation $\Re_x$ on $\mathfrak{L}$ whose matrix is Γ_x is the correspondence

$$a \to aR_x = a \cdot x, \tag{8}$$

and is called a *right multiplication* of $\mathfrak{A}$. Its matrix Γ_x depends upon our choice of the linear space equivalence between $\mathfrak{A}$ and $\mathfrak{L}$. However the transformation R_x does not depend upon this choice. We shall therefore study the properties of R_x rather than of Γ_x. However our present notation aR_x for the result of applying R_x to a has the advantage that $aR_x = a\Gamma_x$, so that in computations we may replace R_x by its matrix. We shall not use the author's earlier notation a^{R_x}.

The set

$$R(\mathfrak{A}) \tag{9}$$

of all right multiplications of $\mathfrak{A}$ is a linear subspace of order at most n of the total matric algebra $(\mathfrak{F})_n$ (of order n^2 over $\mathfrak{F}$) of all linear transformations on $\mathfrak{L}$. The correspondence

$$x \to R_x \tag{10}$$

is a linear mapping of $\mathfrak{L}$ on $R(\mathfrak{A})$, and $R(\mathfrak{A})$ is spanned by $R_{e_1}, \cdots, R_{e_n}$.

We may now state that any algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$ consists of a linear space $\mathfrak{L}$ of this same order, a linear space $R(\mathfrak{A})$ of order $m \leqq n$ over $\mathfrak{F}$ consisting of linear transformations on $\mathfrak{L}$, and a linear mapping (10) of $\mathfrak{L}$ on $R(\mathfrak{A})$. Conversely let $\mathfrak{L}$ and a subspace $\mathfrak{R}$ of $(\mathfrak{F})_n$ be given such that the order of $\mathfrak{R}$ is at most n. Then we may select any n transformations $R^{(i)}$ which span $\mathfrak{R}$ and define $R_x = R^{(1)}\xi_1 + \cdots + R^{(n)}\xi_n$. This then determines a linear mapping of $\mathfrak{L}$ on $\mathfrak{R}$ and hence an algebra $\mathfrak{A}$ with $\mathfrak{R} = R(\mathfrak{A})$. It is particularly important to note that $\mathfrak{R}$ is completely arbitrary save for the upper bound n on its order.

The linear transformations L_x given by

$$a \to x \cdot a = aL_x \tag{11}$$

are called *left multiplications* of $\mathfrak{A}$ and form the *left multiplication space* $L(\mathfrak{A})$ of $\mathfrak{A}$. This space and the linear mapping

$$x \to L_x \tag{12}$$

of $\mathfrak{L}$ on $L(\mathfrak{A})$ determine and are completely determined by $R(\mathfrak{A})$ and (10). For the matrix of L_x is $\Delta_x = \Delta^{(1)}\xi_1 + \cdots + \Delta^{(n)}\xi_n$, where $\Delta^{(i)}$ is the matrix with γ_{ijk} in the j^{th} row and k^{th} column. Correspondingly $L^{(1)}\xi_1 + \cdots + L^{(n)}\xi_n = L_x$ so that $L(\mathfrak{A})$ is spanned over $\mathfrak{F}$ by $L^{(1)}, \cdots, L^{(n)}$.

The right and left multiplication spaces of $\mathfrak{A}$ generate another linear sub-

space of $(\mathfrak{F})_n$ which we shall call the *transformation algebra* of $\mathfrak{A}$. It is the algebra

$$(13) \qquad T(\mathfrak{A}) = \mathfrak{F}[I, R^{(1)}, \cdots, R^{(n)}, L^{(1)}, \cdots, L^{(n)}]$$

of all polynomials with coefficients in $\mathfrak{F}$ in the $R^{(i)}$ which span $R(\mathfrak{A})$, the $L^{(i)}$ which span $L(\mathfrak{A})$ and the identity transformation I of $(\mathfrak{F})_n$. All three spaces $R(\mathfrak{A})$, $L(\mathfrak{A})$, $T(\mathfrak{A})$ will be used to describe properties of $\mathfrak{A}$ and we shall call them the *multiplication spaces* of $\mathfrak{A}$.

The scalar extension $\mathfrak{A}_\mathfrak{K}$ of $\mathfrak{A}$ by any scalar extension field $\mathfrak{K}$ of $\mathfrak{F}$ is the set of vectors (1) with the α_i in $\mathfrak{K}$ and with $a \cdot x$ defined in $\mathfrak{A}_\mathfrak{K}$ by the same γ_{ijk} as define $\mathfrak{A}$. Then clearly we have the same $R^{(i)}$ and $L^{(i)}$, and

$$(14) \qquad R(\mathfrak{A}_\mathfrak{K}) = [R(\mathfrak{A})]_\mathfrak{K}, \quad L(\mathfrak{A}_\mathfrak{K}) = [L(\mathfrak{A})]_\mathfrak{K}, \quad T(\mathfrak{A}_\mathfrak{K}) = [T(\mathfrak{A})]_\mathfrak{K}.$$

When we begin to discuss more than one algebra $\mathfrak{A}$ defined for the same $\mathfrak{L}$ it will be necessary to distinguish $\mathfrak{A}$ from the fixed linear space $\mathfrak{L}$ of the quantities of $\mathfrak{A}$. However no confusion will arise if we speak of a linear subspace of $\mathfrak{L}$ as a *subspace of* $\mathfrak{A}$ and this will be desirable, of course, in discussing subalgebras of $\mathfrak{A}$.

3. Products of spaces

In our study of the multiplication spaces of an algebra we shall need to use the notations for products of spaces of both the algebra $\mathfrak{A}$ and the algebra $(\mathfrak{F})_n$. We define the *product*

$$\mathfrak{B}\mathfrak{C}$$

of any two linear subspaces of an algebra $\mathfrak{A}$ to be the linear subspace over $\mathfrak{F}$ of $\mathfrak{A}$ spanned by $b_i \cdot c_j$ $(i = 1, \cdots, s; j = 1, \cdots, t)$, where $b_1, \cdots, b_s$ span $\mathfrak{B}$ and $c_1, \cdots, c_t$ span $\mathfrak{C}$. Then the square $\mathfrak{B}^{\cdot 2} = \mathfrak{B}\mathfrak{B}$ is defined and we define the right power $\mathfrak{B}^{\cdot k+1} = \mathfrak{B}^{\cdot k}\mathfrak{B}$.

If a is in $\mathfrak{A}$ the subspace (of order zero or one over $\mathfrak{F}$) of $\mathfrak{A}$ spanned by a will be designated by $a\mathfrak{F}$. Then we write $a\mathfrak{B}$ for $(a\mathfrak{F})\mathfrak{B}$ and similarly $\mathfrak{B}a$ for $\mathfrak{B}(a\mathfrak{F})$. If c is also in $\mathfrak{A}$ we write $(a\mathfrak{B})c$ for $(a\mathfrak{B})(c\mathfrak{F})$ and $a(\mathfrak{B}c)$ for $(a\mathfrak{F})(\mathfrak{B}c)$. If $\mathfrak{A}$ is associative and $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{D}$ are linear subspaces then $(\mathfrak{B}\mathfrak{C})\mathfrak{D} = \mathfrak{B}(\mathfrak{C}\mathfrak{D})$, $b\mathfrak{C}d$ is defined to be $(b\mathfrak{C})d = b(\mathfrak{C}d)$ for every b and d in $\mathfrak{A}$.

The definitions above apply of course both to subspaces of any algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$ and to subspaces of $(\mathfrak{F})_n$. However let $\mathfrak{B}$ be a linear subspace of $\mathfrak{A}$ and $\mathfrak{S}$ be a linear subspace of $(\mathfrak{F})_n$. We define

$$(15) \qquad \mathfrak{B}\mathfrak{S}$$

to be the linear subspace of $\mathfrak{A}$ spanned over $\mathfrak{F}$ by the products bS for b in $\mathfrak{B}$ and S in $\mathfrak{S}$. Then we have defined the *product operation* (15) as an operation on $\mathfrak{A}$, $(\mathfrak{F})_n$ to $\mathfrak{A}$. We also define $b\mathfrak{S} = (b\mathfrak{F})\mathfrak{S}$, $\mathfrak{B}S = \mathfrak{B}(S\mathfrak{F})$ for all linear subspaces $\mathfrak{B}$ of $\mathfrak{A}$ and $\mathfrak{S}$ of $(\mathfrak{F})_n$, and all quantities b in $\mathfrak{A}$ and S in $(\mathfrak{F})_n$. Clearly $(\mathfrak{B}\mathfrak{S})\mathfrak{T} = \mathfrak{B}(\mathfrak{S}\mathfrak{T})$ for all linear subspaces $\mathfrak{B}$ of $\mathfrak{A}$ and $\mathfrak{S}$ and $\mathfrak{T}$ of $(\mathfrak{F})_n$. We now prove N. Jacobson's

LEMMA 1. *Let $\mathfrak{N}$ be a nilpotent subalgebra of $(\mathfrak{F})_n$. Then $\mathfrak{A}\mathfrak{N} \neq \mathfrak{A}$ or zero.*

For $\mathfrak{N} \neq 0$ and contains an $S \neq 0$ of $(\mathfrak{F})_n$, $aS \neq 0$ for some a of $\mathfrak{A}$ and is in $\mathfrak{A}\mathfrak{N} \neq 0$. If $\mathfrak{A}\mathfrak{N} = \mathfrak{A}$ then $\mathfrak{A}\mathfrak{N}^{\cdot 2} = (\mathfrak{A}\mathfrak{N})\mathfrak{N} = \mathfrak{A}\mathfrak{N} = \mathfrak{A}$ and $\mathfrak{A}\mathfrak{N}^{\cdot k} = \mathfrak{A}$ implies that $\mathfrak{A}\mathfrak{N}^{\cdot k+1} = (\mathfrak{A}\mathfrak{N}^{\cdot k})\mathfrak{N} = \mathfrak{A}\mathfrak{N} = \mathfrak{A}$. Hence $\mathfrak{A}\mathfrak{N}^{\cdot t} = \mathfrak{A}$ for every t. But $\mathfrak{N}^{\cdot t} = 0$ for some t, $\mathfrak{A} = 0$ which is impossible.

If E is in $(\mathfrak{F})_n$ and $\mathfrak{S}$ is a linear subspace of $(\mathfrak{F})_n$ the condition $E\mathfrak{S} = E\mathfrak{S}E$ means that $ES = EUE$ for every S of $\mathfrak{S}$ where U in $\mathfrak{S}$ is determined (but not necessarily uniquely) by S. However when E is an idempotent, that is $E^2 = E$, the property $E\mathfrak{S} = E\mathfrak{S}E$ is equivalent to $ES = ESE$ for every S of $\mathfrak{S}$. For $ES = EUE = EUE^2 = (EUE)E = ESE$.

4. Subalgebras

If $\mathfrak{B}$ is a linear subspace of an algebra $\mathfrak{A}$ the set of all R_y for y in $\mathfrak{B}$ is a linear subspace of $R(\mathfrak{A})$, the set of all L_y is a linear subspace of $L(\mathfrak{A})$, and these subspaces, together with I, generate a subalgebra of $T(\mathfrak{A})$. We designate these three linear subspaces of $T(\mathfrak{A})$ by

$$R(\mathfrak{B}, \mathfrak{A}), \qquad L(\mathfrak{B}, \mathfrak{A}), \qquad T(\mathfrak{B}, \mathfrak{A}) \tag{16}$$

respectively, where $T(\mathfrak{B}, \mathfrak{A})$ is the set of all polynomials with coefficients in $\mathfrak{F}$ in the R_y, the L_y and I.

If $\mathfrak{B}$ has order $m < n$ over $\mathfrak{F}$ we may express $\mathfrak{A}$ as the supplementary sum $\mathfrak{B} + \mathfrak{C}$ where $\mathfrak{C}$ has order $n - m$. This means that every a of $\mathfrak{A}$ is uniquely expressible in the form $a = b + c$ for b in $\mathfrak{B}$ and c in $\mathfrak{C}$. However $\mathfrak{C}$ is by no means unique. We now define a mapping

$$E: \qquad a = b + c \rightarrow b = aE$$

of $\mathfrak{A}$ on $\mathfrak{B}$. It is an idempotent linear transformation of rank m, that is, $E^2 = E$ and m is the rank of the matrix of E. We then have $\mathfrak{B} = \mathfrak{A}E$ where E is characterized by the property that $a = aE$ if and only if a is in $\mathfrak{B}$, $aE = 0$ if and only if a is in $\mathfrak{C}$. A corresponding idempotent for $\mathfrak{C}$ is $I - E$. We now have

LEMMA 2. *Let $\mathfrak{B}$ be a linear subspace of order m of $\mathfrak{A}$ so that $\mathfrak{B} = \mathfrak{A}E$ for an idempotent E of rank m in $(\mathfrak{F})_n$. Then $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}$ if and only if* $E[R(\mathfrak{B}, \mathfrak{A})] = E[R(\mathfrak{B}, \mathfrak{A})]E$.

For $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}$ if and only if $aE \cdot y = (aE \cdot y)E$ for every a of $\mathfrak{A}$ and y of B. Then $aER_y = aER_yE$, $ER_y = ER_yE$ and we have our lemma since $E^2 = E$.

Note that also $y \cdot aE = aEL_y = aEL_yE$, $E[L(\mathfrak{B}, \mathfrak{A})] = E[L(\mathfrak{B}, \mathfrak{A})]E$. Since $EIE = E = EI$ we see that $EU = EUE$ for every U of IF, $R(\mathfrak{B}, \mathfrak{A})$, $L(\mathfrak{B}, \mathfrak{A})$. But if $ES = ESE$ and $EU = EUE$ we have $E(S + U) = E(S + U)E$, $ESU = ESEU = ESEUE = ESUE$. Hence $E[T(\mathfrak{B}, \mathfrak{A})] = E[T(\mathfrak{B}, \mathfrak{A})]E$. The converse is trivial and we have

LEMMA 2′. *Let $\mathfrak{B} = \mathfrak{A}E$ as in Lemma 2. Then $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}$ if and only if* $E[T(\mathfrak{B}, \mathfrak{A})] = E[T(\mathfrak{B}, \mathfrak{A})]E$.

5. Ideals

A subspace $\mathfrak{B}$ of an algebra $\mathfrak{A}$ is a right ideal of $\mathfrak{A}$ if and only if $y \cdot x$ is in $\mathfrak{B}$ for every y of $\mathfrak{B}$ and x of $\mathfrak{A}$. Then $\mathfrak{B} = \mathfrak{A}E$ for an idempotent E of rank equal to the order of $\mathfrak{B}$ over $\mathfrak{F}$ and $\mathfrak{B}$ is a right ideal of $\mathfrak{A}$ if and only if either of the following conditions

$$(17) \qquad L_y = L_y E, \qquad ER_x = ER_x E \qquad (x \text{ in } \mathfrak{A},\ y = yE \text{ in } \mathfrak{B})$$

holds. For $y \cdot x = xL_y = xL_yE$, $y = aE$, $aE \cdot x = aER_x = aER_xE$. We may state this result as

LEMMA 3. *Let $\mathfrak{B} = \mathfrak{A}E$ for an idempotent E of $\mathfrak{A}$. Then $\mathfrak{B}$ is a right ideal of $\mathfrak{A}$ if and only if $ER(\mathfrak{A}) = ER(\mathfrak{A})E$. This is equivalent to the condition that $L(\mathfrak{B}, \mathfrak{A})$ be contained in $[L(\mathfrak{A})]E$.*

In the theory of group representations the property $ER(\mathfrak{A}) = ER(\mathfrak{A})E$ for $E \neq 0, I$ is called the property that $R(\mathfrak{A})$ is a *reducible set* of linear transformations. We shall not use this terminology again here.

Left ideals are defined similarly and $\mathfrak{B} = \mathfrak{A}E$ is a left ideal if and only if $EL(\mathfrak{A}) = EL(\mathfrak{A})E$, and thus if and only if $R(\mathfrak{B}, \mathfrak{A})$ is in $[R(\mathfrak{A})]E$. We call $\mathfrak{B}$ an ideal of $\mathfrak{A}$ if it is both a left and a right ideal. This occurs if and only if $\mathfrak{B} = \mathfrak{A}E$ where $EU = EUE$ for every U in either $R(\mathfrak{A})$ or $L(\mathfrak{A})$. As in the proof of Lemma 2 we have $EU = EUE$ for every U of $T(\mathfrak{A})$ and have

LEMMA 4. *A linear subspace $\mathfrak{B} = \mathfrak{A}E$ of $\mathfrak{A}$ is an ideal of $\mathfrak{A}$ if and only if $ET(\mathfrak{A}) = ET(\mathfrak{A})E$.*

We shall call a quantity a of $\mathfrak{A}$ *right singular* or *right non-singular* according as R_a is or is not singular. We then have

LEMMA 5. *Let $\mathfrak{B} = \mathfrak{A}E$ be an ideal of $\mathfrak{A}$ and a be a right non-singular quantity of $\mathfrak{A}$. Then $E(R_a)^{-1} = E(R_a)^{-1}E$.*

For R_a is in the associative algebra $T(\mathfrak{A})$ and so is $(R_a)^{-1}$, $ET(\mathfrak{A}) = ET(\mathfrak{A})E$.

We next prove

LEMMA 6. *Let P be in $(\mathfrak{F})_n$ and $\mathfrak{F}[P]$ be a field of degree n over $\mathfrak{F}$. Then $EP = EPE$ for an idempotent E of $(\mathfrak{F})_n$ if and only if $E = I$ or $E = 0$.*

For let $EP = EPE$ and $E \neq 0$. Then EP is in the total matric algebra $E(\mathfrak{F})_nE$ with unity quantity E, a total matric algebra whose degree m is the rank of E. If $EP^k = EP^kE$ then $EP^{k+1} = EP^kEP = EP^kEPE = EP^{k+1}E$, $EP^t = EP^tE$ for every t. But then $(EP)^t = EP^tE$ since from $(EP)^k = EP^kE$ we have $(EP)^{k+1} = EP^kEEP = EP^{k+1}E$. It follows that $\phi(EP) = E\phi(P)E$ for any polynomial $\phi(P)$. But if $\phi(\lambda)$ is the minimum function of P it is an irreducible polynomial of degree n and $\phi(EP) = 0$. This is impossible when E is singular since the minimum function of EP in a total matric algebra of degree $m < n$ has degree at most m. Thus $E = I$.

6. Divisors of zero

If b is right non-singular the equation $x \cdot b = a$ has the unique solution $x = a(R_b)^{-1}$. However there exists a $c \neq 0$ such that $c \cdot b = 0$ when b is right singular. We shall call b a *right divisor of zero* if it is a non-zero right singular

quantity. *Left singularity* and *left non-singularity* as well as *left divisors of zero* are defined similarly, and it is clear that an algebra contains right divisors of zero b if and only if it contains left divisors of zero c.

A quantity b of an algebra $\mathfrak{A}$ is called an *absolute right divisor of zero* if $b \neq 0$ and $a \cdot b = 0$ for every a of $\mathfrak{A}$. But then $R_b = 0$. This can occur only if the linear mapping $x \to R_x$ of $\mathfrak{A}$ on $R(\mathfrak{A})$ is singular, that is, if and only if $R(\mathfrak{A})$ has smaller order than $\mathfrak{A}$. Similarly $L(\mathfrak{A})$ has order less than the order of $\mathfrak{A}$ if and only if some quantity b in $\mathfrak{A}$ is an absolute left divisor of zero, that is, $L_b = 0$ and $b \neq 0$. Each absolute right (left) divisor of zero spans a subalgebra $\mathfrak{B} = b\mathfrak{F}$ of $\mathfrak{A}$ which is a zero algebra of order one and is a left (right) ideal of $\mathfrak{A}$. In fact we have

LEMMA 7. *A linear subspace $\mathfrak{B} = \mathfrak{A}E = b\mathfrak{F}$ for an absolute right divisor of zero b of $\mathfrak{A}$ if and only if $EL(\mathfrak{A}) = 0$, E has rank one.*

For $\mathfrak{B} = b\mathfrak{F} = \mathfrak{A}E$ where E has rank one. Then aE is zero or an absolute right divisor of zero if and only if $x \cdot aE = aEL_x = 0$, $EL_x = 0$, $EL(\mathfrak{A}) = 0$.

If $T(\mathfrak{A}) = I\mathfrak{F}$ then $R_a = \alpha I$, $L_a = \beta I$ for every a of $\mathfrak{A}$ where α and β are in $\mathfrak{F}$. If $R_a \neq 0$ for some a then $x \cdot a = aL_x = xR_a = x\alpha \neq 0$ and $L_x \neq 0$. It follows that the mapping $x \to L_x$ is one-to-one, $n = 1$. Similarly if $L_a \neq 0$ we have $n = 1$. Thus $n > 1$ implies that $T(\mathfrak{A}) = I\mathfrak{F}$ only if $R(\mathfrak{A}) = L(\mathfrak{A}) = 0$, $\mathfrak{A}$ is a zero algebra. The converse is trivial and we have

LEMMA 8. *The algebra $T(\mathfrak{A}) = I\mathfrak{F}$ if and only if $\mathfrak{A}$ is a zero algebra or $n = 1$ and $\mathfrak{A} = \mathfrak{F}$.*

We call b an *absolute divisor of zero* if it is both an absolute right and an absolute left divisor of zero. For algebras containing such quantities we have

LEMMA 9. *An algebra $\mathfrak{A}$ contains absolute divisors of zero if and only if there exists a non-zero idempotent E of $(\mathfrak{F})_n$ such that $ER(\mathfrak{A}) = EL(\mathfrak{A}) = 0$. Then*

$$T(\mathfrak{A}) = \mathfrak{S} + I\mathfrak{F} \tag{18}$$

where $\mathfrak{S}$ is the set of all transformations S of $T(\mathfrak{A})$ such that $ES = 0$ and is an ideal of $T(\mathfrak{A})$.

For if b is an absolute divisor of zero the proof of Lemma 7 implies that there exists a non-zero idempotent E such that $ER(\mathfrak{A}) = EL(\mathfrak{A}) = 0$. Conversely if $ER(\mathfrak{A}) = EL(\mathfrak{A}) = 0$ the quantities of $\mathfrak{A}E$ have the property $aE \cdot x = aER_x = 0 = aEL_x = x \cdot aE$. Then $\mathfrak{A}E \neq 0$ consists of zero and absolute divisors of zero. The quantities of $T(\mathfrak{A})$ are sums of scalars αI for α in $\mathfrak{F}$ and products $U = U_1 \cdots U_t$ for the U_i in $R(\mathfrak{A})$ and $L(\mathfrak{A})$. Then $EU = 0$. Hence every transformation of $T(\mathfrak{A})$ is expressible as a sum $S + \alpha I$ with $ES = 0$ and α in $\mathfrak{F}$. Since $T(\mathfrak{A})$ contains $I\mathfrak{F}$ we have S in $T(\mathfrak{A})$, and (18) holds. That $\mathfrak{S}$ is an ideal of $\mathfrak{A}$ follows from the property that if U and S are in $\mathfrak{S}$ then $E[U(S + \alpha I)] = EU(S + \alpha I) = 0$, $E[(S + \alpha I)U] = ESU + EU\alpha = 0$.

7. Simple algebras

An algebra $\mathfrak{A}$ is said to be *simple* if $\mathfrak{A}$ is not a zero algebra of order one and $\mathfrak{A}$ is the only non-zero ideal of $\mathfrak{A}$. We define the $\mathfrak{A}$-*centralizer* of any set $\mathfrak{H}$ of

quantities of any algebra $\mathfrak{A}$ to be the set of all quantities k in $\mathfrak{A}$ such that $k \cdot h = h \cdot k$ for every h of $\mathfrak{H}$ and see that this set is a subalgebra of $\mathfrak{A}$ if $\mathfrak{A}$ is associative. Then we may prove

LEMMA 10. *The algebra $T(\mathfrak{A})$ is simple if and only if $\mathfrak{A}$ is either simple or $T(\mathfrak{A}) = I\mathfrak{F}$ and $\mathfrak{A}$ is a zero algebra.*

For let $T(\mathfrak{A})$ be simple. By Lemma 9 if $\mathfrak{A}$ contains an absolute divisor of zero we have $\mathfrak{S} = 0$ in (18) and $\mathfrak{A}$ is a zero algebra, by Lemma 8. Hence let $\mathfrak{A}$ be not a zero algebra and suppose that $\mathfrak{B} = \mathfrak{A}E$ is a non-zero ideal of $\mathfrak{A}$ for an idempotent $E \neq 0$ of $(\mathfrak{F})_n$. We define $\mathfrak{S}$ to be the set of all S in $T(\mathfrak{A})$ such that $S = \mathfrak{S}E$. By (17) $\mathfrak{S}$ contains $L(\mathfrak{B}, \mathfrak{A})$ and similarly contains $R(\mathfrak{B}, \mathfrak{A})$. But if $y \neq 0$ and $L_y = 0$ we have $R_y \neq 0$ since $\mathfrak{A}$ contains no absolute divisor of zero. Hence $L(\mathfrak{B}, \mathfrak{A})$ and $R(\mathfrak{B}, \mathfrak{A})$ are not both zero, $\mathfrak{S} \neq 0$. If S is in $\mathfrak{S}$ and U is in $T(\mathfrak{A})$ we have $SU = SEU = SEUE = (SU)E$ by Lemma 4 while also $US = (US)E$. But then SU and US are in $\mathfrak{S}$, $\mathfrak{S}$ is an ideal of $T(\mathfrak{A})$, $\mathfrak{S} = T(\mathfrak{A})$ contains $I = IE = E$, $\mathfrak{B} = \mathfrak{A}E = \mathfrak{A}$ is simple.

Conversely let $\mathfrak{A}$ be simple. If $T(\mathfrak{A})$ has a nilpotent ideal $\mathfrak{N}$ we have $\mathfrak{N}R(\mathfrak{A})$ in $\mathfrak{N}$, $\mathfrak{N}L(\mathfrak{A})$ in $\mathfrak{N}$ so that if $\mathfrak{B} = \mathfrak{A}\mathfrak{N}$ we have $\mathfrak{B}\mathfrak{A} = \mathfrak{B}R(\mathfrak{A}) = \mathfrak{A}\mathfrak{N}R(\mathfrak{A})$ contained in $\mathfrak{B}$, $\mathfrak{A}\mathfrak{B} = \mathfrak{B}L(\mathfrak{A}) = \mathfrak{A}\mathfrak{N}L(\mathfrak{A})$ in $\mathfrak{B}$. Hence $\mathfrak{B}$ is an ideal of $\mathfrak{A}$. But by Lemma 1 $\mathfrak{B} \neq 0$, $\mathfrak{A}$ contrary to hypothesis. Hence $T(\mathfrak{A})$ is an associative semi-simple algebra and is either simple as desired or is a direct sum. In the latter case the unity quantity of a component of $T(\mathfrak{A})$ is an idempotent E, in the $(\mathfrak{F})_n$-centralizer of $T(\mathfrak{A})$, which is singular and not zero. Then by Lemma 4 $\mathfrak{B} = \mathfrak{A}E$ is an ideal of $\mathfrak{A}$, $\mathfrak{B} \neq 0$ or $\mathfrak{A}$. This completes our proof.

8. Central simple algebras

A field $\mathfrak{C}$ consisting of linear transformations over $\mathfrak{F}$ on a linear space $\mathfrak{A}$ of order n over $\mathfrak{F}$ is called a *subfield* of $(\mathfrak{F})_n$ if the identity transformation of $(\mathfrak{F})_n$ is in $\mathfrak{C}$. Then $\mathfrak{A}$ may be regarded as being a linear space of order σ over $\mathfrak{C}$ and $n = \sigma\tau$ where τ is the degree of $\mathfrak{C}$ over $\mathfrak{F}$. The set of all linear transformations over $\mathfrak{C}$ on $\mathfrak{A}$ is the total matric algebra $(\mathfrak{C})_\sigma$ and is clearly the $(\mathfrak{F})_n$-centralizer of $\mathfrak{C}$. The equations $a \cdot x = aR_x$, $x \cdot a = aL_x$ then define the algebra $\mathfrak{A}$ over $\mathfrak{F}$ as an algebra over $\mathfrak{C}$ if and only if every R_x and L_x is in $(\mathfrak{C})_\sigma$. But then $R(\mathfrak{A})$, $L(\mathfrak{A})$, $T(\mathfrak{A})$ are in $(\mathfrak{C})_\sigma$. It follows that $\mathfrak{A}$ is an algebra over a subfield $\mathfrak{C}$ of $(\mathfrak{F})_n$ if and only if $\mathfrak{C}$ is in the $(\mathfrak{F})_n$-centralizer of $T(\mathfrak{A})$.

We define the *transformation center* of $\mathfrak{A}$ to be the $T(\mathfrak{A})$-centralizer of $T(\mathfrak{A})$ and designate this subalgebra of $T(\mathfrak{A})$ by $\mathfrak{C}(\mathfrak{A})$. If $T(\mathfrak{A})$ is simple the transformation center of $\mathfrak{A}$ is a field of degree t over $\mathfrak{F}$ and $n = st$, $T(\mathfrak{A})$ is contained in the total matric algebra $[\mathfrak{C}(\mathfrak{A})]_s$.

An algebra $\mathfrak{A}$ over $\mathfrak{F}$ is said to be *central simple over* $\mathfrak{F}$ if $\mathfrak{A}_\mathfrak{K}$ is simple for every scalr extension $\mathfrak{K}$ of $\mathfrak{F}$. If then $\mathfrak{A}$ is simple and $\mathfrak{Z}$ is any subfield of its transformation center $\mathfrak{C} = \mathfrak{C}(\mathfrak{A})$ the degree of $\mathfrak{Z}$ divides $t = \tau\rho$ and $\mathfrak{A}$ is simple of order $s\rho$ over $\mathfrak{Z}$, $T(\mathfrak{A})$ is simple over $\mathfrak{Z}$ and is contained in $\mathfrak{Z}_{s\rho}$. But $[T(\mathfrak{A})]_\mathfrak{K} = T(\mathfrak{A}_\mathfrak{K})$ for every scalar extension $\mathfrak{K}$ of $\mathfrak{Z}$ and thus $\mathfrak{A}$ is central simple over $\mathfrak{Z}$

only if $T(\mathfrak{A})$ is central simple over $\mathfrak{Z}$. This can occur only if $\mathfrak{Z} = \mathfrak{C}(\mathfrak{A})$. We use this result and then prove[6]

THEOREM 1. *An algebra $\mathfrak{A}$ of order $n > 1$ over $\mathfrak{F}$ is simple if and only if $T(\mathfrak{A})$ is the total matric algebra $(\mathfrak{C})_s$ where $\mathfrak{C} = \mathfrak{C}(\mathfrak{A})$ is a field of degree t over $\mathfrak{F}$ and $n = st$. Moreover $\mathfrak{A}$ is central simple over a subfield $\mathfrak{Z}$ of $(\mathfrak{F})_n$ if and only if $\mathfrak{Z} = \mathfrak{C}$.*

For if $\mathfrak{A}$ is simple so is $\mathfrak{A}$ over its transformation center $\mathfrak{C}$ and $\mathfrak{A}_\mathfrak{K}$ is not a zero algebra for any scalar extension $\mathfrak{K}$ of $\mathfrak{C}$. Now $T(\mathfrak{A})$ is in $(\mathfrak{C})_s$ and is known[7] to be a central simple algebra over $\mathfrak{C}$. The $(\mathfrak{C})_s$-centralizer of $T(\mathfrak{A})$ is also a central simple algebra $\mathfrak{D}$ of degree q over $\mathfrak{C}$ and $T(\mathfrak{A}) = (\mathfrak{C})_s$ if and only if $q = 1$. If $q > 1$ we let $\mathfrak{K}$ be a splitting field over $\mathfrak{C}$ of $\mathfrak{D}$ and see that the total matric algebra $\mathfrak{D}_\mathfrak{K}$ contains a non-zero idempotent E which is singular and in the $(\mathfrak{C})_s$-centralizer of $T(\mathfrak{A}_\mathfrak{K})$. Then by Lemma 4 $\mathfrak{A}_\mathfrak{K}E$ is a non-zero proper ideal over $\mathfrak{K}$ of $\mathfrak{A}_\mathfrak{K}$, whereas the proof above shows that $\mathfrak{A}_\mathfrak{K}$ is simple, a contradiction. It follows that $T(\mathfrak{A}) = (\mathfrak{C})_s$. The only subfields $\mathfrak{Z}$ of $(\mathfrak{F})_n$ in the $(\mathfrak{F})_n$-centralizer of $T(\mathfrak{A})$ are in the $(\mathfrak{F})_n$-centralizer of $\mathfrak{C}$ and hence in $(\mathfrak{C})_s$. They are then in the $(\mathfrak{C})_s$-centralizer of $(\mathfrak{C})_s$ and thus in $\mathfrak{C}$, $\mathfrak{A}$ is central simple over $\mathfrak{Z}$ only if $\mathfrak{Z} = \mathfrak{C}$. Conversely let $T(\mathfrak{A}) = (\mathfrak{C})_s$ where $n = st$ and $(\mathfrak{C})_s$ is a total matric algebra of degree s over $\mathfrak{C}$, $\mathfrak{C}$ is a field of degree t over $\mathfrak{F}$. Then the order of $T(\mathfrak{A})$ is $s^2t > 1$ since otherwise $s = t = n = 1$. Hence $\mathfrak{A}$ is not a zero algebra and, by Lemma 9, $\mathfrak{A}$ is simple. Also $\mathfrak{A}$ is central simple over $\mathfrak{C}$ since $T(\mathfrak{A})$ is a total matric algebra over $\mathfrak{C}$ and is central simple, $\mathfrak{A}_\mathfrak{K}$ is not a zero algebra over any scalar extension $\mathfrak{K}$ of $\mathfrak{C}$, $\mathfrak{A}_\mathfrak{K}$ is simple. Thus $\mathfrak{A}$ is central simple over its transformation center.

9. Algebras with a left unity quantity

Let e be a non-zero vector in a linear space of order n over $\mathfrak{F}$ and $\mathfrak{S}$ be any linear subspace of order $m \leqq n$ of $(\mathfrak{F})_n$. Then $e\mathfrak{S}$ is a linear subspace of $\mathfrak{L}$ and the correspondence

$$S \rightarrow eS \qquad (S \text{ in } \mathfrak{S}), \tag{19}$$

is a linear mapping of $\mathfrak{S}$ on $e\mathfrak{S}$. It follows that the order of $e\mathfrak{S}$ over $\mathfrak{F}$ is at most m.

---

[6] Results essentially equivalent to this one and to Lemma 10 were given by N. Jacobson, *A note on non-associative algebras*, Duke J., vol. 3 (1937), pp. 544–8. The result was first announced for Lie algebras by the author in the A. M. S. Bulletin, vol. 41 (1935), p. 344, and the author feels that the present exposition is not only in a form better suited than that of Jacobson for later application but presents also a much clearer picture of the relations between an algebra $\mathfrak{A}$ and its transformation algebra $T(\mathfrak{A})$. The idea of studying these relations was suggested to both Jacobson and the author by the lectures of H. Weyl on Lie Algebras which were given in Fine Hall in 1933. Note that if $\mathfrak{T}$ is the algebra generated by the right and left multiplications of $\mathfrak{A}$ then $\mathfrak{T} = T(\mathfrak{A})$ unless $\mathfrak{T}$ does not contain the identity transformation. But then $\mathfrak{T}$ is an ideal of $T(\mathfrak{A})$ and this cannot occur when $\mathfrak{A}$ is simple.

[7] For the properties used here see Chapters I, III, IV of the author's *Structure of Algebras*.

Suppose then that (19) is one-to-one, and that $m = n$. Then (19) maps $\mathfrak{S}$ on $\mathfrak{L}$ such that $x = eS = eT$ if and only if $S = T$. We may then define R_x as that transformation S for which $eS = x$ and have defined an algebra $\mathfrak{A}$ without absolute right divisors of zero and such that $\mathfrak{S} = R(\mathfrak{A})$, $e \cdot x = eR_x = x$ for every x of $\mathfrak{A}$.

The quantity e is now a left unity quantity of $\mathfrak{A}$. Conversely every algebra $\mathfrak{A}$ with a left unity quantity e has no absolute right divisors of zero and is such that the linear mapping $R_x \to x = eR_x$ is a one-to-one mapping of $R(\mathfrak{A})$ on $\mathfrak{A} = eR(\mathfrak{A})$. Then multiplication in $\mathfrak{A}$ is given by

$$eS \cdot eR = eSR \tag{20}$$

for every S of $(\mathfrak{F})_n$ and every R of $R(\mathfrak{A})$. It may be seen, however, that (20) does not hold if R is not in $R(\mathfrak{A})$.

If e is given we say that b in $\mathfrak{A}$ is *right regular with respect to* e if $c \cdot b = e$ for c in $\mathfrak{A}$. Then c is a *left inverse* of b (relative to e). *Such a quantity may exist even when* b *is right singular.*

The left inverse of a right non-singular quantity b may be expressed as a certain polynomial in b. We define $b^2 = bb$ and then the right powers of b by $b^{k+1} = (b^k)b$ for $k > 0$, $b^k = e$ for $k = 0$. If λ is an indeterminate over $\mathfrak{F}$ and

$$\phi(\lambda) = \lambda^t + \beta_1\lambda^{t-1} + \cdots + \beta_t \qquad (\beta_i \text{ in } \mathfrak{F}) \tag{21}$$

we define the right polynomial

$$\phi_R(b) = b^t + \beta_1 b^{t-1} + \cdots + \beta_{t-1}b + \beta_t e \tag{22}$$

for any b in $\mathfrak{A}$ and right powers b^k. Then it is clear from (26) that

$$\phi_R(b) = e\phi(R_b). \tag{23}$$

Hence $\phi_R(b) = 0$ if $\phi(R_b) = 0$. However $\phi_R(b)$ may be zero for $\phi(R_b)$ a non-zero linear transformation carrying the vector e into zero. Note now that in particular

$$f_R(b) = g_R(b) = 0 \tag{24}$$

where $f(\lambda)$ is the characteristic function and $g(\lambda)$ is the minimum function of R_b.

We define the *right minimum function* (with respect to e) of a quantity b of $\mathfrak{A}$ to be the polynomial (21) of least degree t such that $\phi_R(b) = 0$. Its uniqueness is then implied by

THEOREM 2. *The right minimum function* $\phi(\lambda)$ *of a quantity* b *of an algebra* $\mathfrak{A}$ *divides every* $\psi(\lambda)$ *such that* $\psi_R(b) = 0$.

For we write $\psi(\lambda) = \phi(\lambda)\rho(\lambda) + \sigma(\lambda)$ and have $\psi_R(b) = e\psi(R_b) = e[\phi(R_b)\rho(R_b) + \sigma(R_b)] = e\sigma(R_b) = \sigma_R(b)$. But the degree of $\sigma(\lambda)$ may be taken to be less than that of $\phi(\lambda)$, $\sigma(\lambda)$ is identically zero.

We see in particular that the right minimum function $\phi(\lambda)$ of b divides the minimum and characteristic functions of R_b. If b is right non-singular the

constant term of these latter functions is not zero and $\phi(\lambda)$ has the form (27) for $\beta_t = 0$. But then $\phi_R(b) = (b^{t-1} + \beta_1 b^{t-2} + \cdots + \beta_{t-1}e)\cdot b + \beta_t e = 0$,

$$b^{-1}\cdot b = e, \qquad b^{-1} = -\beta_t^{-1}(b^{t-1} + \beta_1 b^{t-2} + \cdots + \beta_{t-1}e). \tag{25}$$

Moreover if $c\cdot b = e$ we have $(c - b^{-1})\cdot b = (c - b^{-1})R_b = 0$ if and only if $c = b^{-1}$.

Right singular quantities may be right regular and it may happen that, while (25) holds for $\beta_t \neq 0$, R_b may be singular. To illustrate this we consider the linear space $\mathfrak{L}$ of order three over a field $\mathfrak{F}$ of characteristic not two where $\mathfrak{L}$ consists of all two-rowed symmetric matrices. Then a basis of $\mathfrak{L}$ is given by

$$e = u_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad u_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad u_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{26}$$

We define an algebra $\mathfrak{A}$ by

$$a\cdot x = \frac{ax + xa}{2} \tag{27}$$

where products on the right are ordinary two-rowed square matrix products. Then $\mathfrak{A}$ is a commutative algebra such that $e\cdot x = x\cdot e = \frac{1}{2}(ex + xe) = x$. But also $x\cdot x = x^2$ and thus

$$u_2\cdot u_2 = u_3\cdot u_3 = e, \tag{28}$$

while

$$2u_2\cdot u_3 = 2u_3\cdot u_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = 0. \tag{29}$$

It follows that u_2 and u_3 are right and left divisors of zero and are right (and left) regular. The (right) minimum function of both u_2 and u_3 is $\lambda^2 - 1$ and the minimum function of R_{u_2} and R_{u_3} is $\lambda(\lambda^2 - 1)$. Here the matrices of these linear transformations with respect to the basis (32) are

$$\Gamma_{u_2} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \Gamma_{u_3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

10. Algebras with a unity quantity

A quantity f of $\mathfrak{A}$ is a *right unity quantity* of $\mathfrak{A}$ if $x\cdot f = x$ for every x of $\mathfrak{A}$, that is, $R_f = I$. Thus the mapping

$$L_x \to fL_x = x \tag{30}$$

of $L(\mathfrak{A})$ on $\mathfrak{A}$ is one-to-one, and multiplication in $\mathfrak{A}$ is defined by

$$eL\cdot eS = eSL \tag{31}$$

for every S of $(\mathfrak{F})_n$ and every L of $R(\mathfrak{A})$. *Left regularity* and *left polynomials* are defined in the obvious fashion and every left non-singular quantity b has a right inverse which is a left polynomial in b.

If $\mathfrak{A}$ has both a left unity quantity e and a right unity quantity f we have both $e \cdot f = f$ and $e \cdot f = e$ so that $e = f$. Then e is the unique quantity of $\mathfrak{A}$ such that $e \cdot x = x \cdot e$ for every x of $\mathfrak{A}$ and we shall call e the *unity quantity* of $\mathfrak{A}$. It has the determining property

$$R_e = L_e = I, \tag{32}$$

and multiplication in $\mathfrak{A}$ is defined by

$$eS \cdot eR = eSR, \qquad eL \cdot eS = eSL \tag{33}$$

for every S of $(\mathfrak{F})_n$, R of $R(\mathfrak{A})$, L of $L(\mathfrak{A})$. Note that then

$$eL \cdot eR = eRL = eLR, \tag{34}$$

so that $e(RL - LR) = 0$ for every R of $R(\mathfrak{A})$ and L of $L(\mathfrak{A})$. However we shall see that $RL = LR$ for every such R and L if and only if $\mathfrak{A}$ is associative.

11. Isotopes of algebras

All algebras $\mathfrak{A}$ of the same order n may be regarded as having quantities comprising the same linear space L of order n over $\mathfrak{F}$. If $\mathfrak{A}$ is given the space $R(\mathfrak{A})$ and the mapping $x \to R_x$ of L on $R(\mathfrak{A})$ are thereby determined and conversely. Thus if $\mathfrak{A}_0$ is a second algebra we have a corresponding linear mapping $x \to R_x^{(0)}$ of $\mathfrak{A}_0$ on a linear subspace $R(\mathfrak{A}_0)$ of $(\mathfrak{F})_n$ and we write

$$(a, x) = aR_x^{(0)} \tag{35}$$

for products in $\mathfrak{A}_0$. We shall now say that $\mathfrak{A}$ *is isotopic to* $\mathfrak{A}_0$ *if there exist non-singular linear transformations* P, Q, C *such that*

$$R_x^{(0)} = PR_{xQ}C, \tag{36}$$

and shall call (36) an *isotopy* of $\mathfrak{A}$ and $\mathfrak{A}_0$.

If $\mathfrak{A}$ is isotopic to $\mathfrak{A}_0$ then $R_{xQ^{-1}}^{(0)} = PR_xC$ so that $R_x = P^{-1}R_{xQ^{-1}}C^{-1}$ and $\mathfrak{A}_0$ is isotopic to $\mathfrak{A}$. Also $\mathfrak{A}$ is isotopic to itself under (36) with $P = Q = C = I$. Finally if $R_x^{(1)} = P_1R_{xQ_1}^{(0)}C_1$ then $R_x^{(1)} = P_2R_{xQ_2}C_2$ where $P_2 = P_1P$, $Q_2 = Q_1Q$, $C_2 = CC_1$. Hence the relation of isotopy is a formal equivalence relation and we shall say that $\mathfrak{A}$ and $\mathfrak{A}_0$ are isotopic as well as that $\mathfrak{A}$ is isotopic to $\mathfrak{A}_0$.

All left multiplications $L_x^{(0)}$ of $\mathfrak{A}_0$ are determined when its right multiplications are given and conversely. Thus we shall determine the conditions relating $L_x^{(0)}$ and L_x which are equivalent to (36). We observe that in $\mathfrak{A}$ we have

$$a \cdot x = aR_x = xL_a, \qquad x \cdot a = aL_x = xR_a, \tag{37}$$

and in $\mathfrak{A}_0$ we have

$$(a, x) = aR_x^{(0)} = xL_a^{(0)}, \qquad (x, a) = aL_x^{(0)} = xR_a^{(0)}. \tag{38}$$

Define

$$b = aQ, \qquad z = xP \tag{39}$$

and obtain

$$aL_x^{(0)} = xR_a^{(0)} = xPR_bC = zR_bC = (z\cdot b)C = bL_zC. \tag{40}$$

Then $aL_x^{(0)} = aQL_zC$ for every a and x of $\mathfrak{L}$ and we have

THEOREM 3. *The conditions* (36) *which imply that* $\mathfrak{A}$ *and* $\mathfrak{A}_0$ *are isotopic are equivalent to*

$$L_x^{(0)} = QL_{xP}C. \tag{41}$$

The relation of equivalence is an instance of isotopy. For two algebras $\mathfrak{A}_0$ and $\mathfrak{A}$ are said to be equivalent if there exists a non-singular linear transformation $a \to aH$ on $\mathfrak{A}_0$ to $\mathfrak{A}$ which is preserved under multiplication. But then

$$(a, x)H = aH\cdot xH, \tag{42}$$

that is $aR_x^{(0)}H = aHR_{xH}$. Hence $\mathfrak{A}_0$ and $\mathfrak{A}$ are equivalent if and only if

$$R_x^{(0)} = HR_{xH}H^{-1}. \tag{43}$$

By Theorem 3 the equivalent algebras $\mathfrak{A}_0$ and $\mathfrak{A}$ are also related by

$$L_x^{(0)} = HL_{xH}H^{-1}. \tag{44}$$

It is usually more convenient to use simplifications of (36) and (41) obtainable by replacing $\mathfrak{A}_0$ by an equivalent algebra. Thus we may apply (43) to (36) with $H = Q^{-1}$ and have $R_x^{(1)} = HR_{xH}^{(0)}H^{-1} = (HP)R_xCH^{-1}$. This result together with Theorem 3 may be stated as

THEOREM 4. *Every isotope of an algebra* $\mathfrak{A}$ *is equivalent to an isotope defined by*

$$R_x^{(0)} = PR_xC, \qquad L_x^{(0)} = L_{xP}C, \tag{45}$$

for non-singular linear transformations P *and* C.

The form above[8] has the advantage that in $\mathfrak{A}_0$ we map x on PR_xC but is so unsymmetrical that we shall prefer the *principal isotopy* obtained from (36) by the application of (43) with $H = C$. We state the result as

---

[8] The concept of isotopy was suggested to the author by the work of N. Steenrod who, in his study of homotopy groups in topology, was led to study isotopy of division algebras. He concluded that algebras related as in (45) would yield the same homotopy properties and should therefore be put in the same class. The author then formulated the concept generally as in (36) and obtained Theorem 3 giving the corresponding property for left multiplications and Theorem 4 showing that Steenrod's isotopes were actually equivalent to the more general type. However the principal isotopes of (46) are much more conveniently handled.

THEOREM 5. *Every isotope of an algebra $\mathfrak{A}$ is equivalent to a* **principal** *isotope $\mathfrak{A}_0$, that is, an isotope with*

$$R_x^{(0)} = PR_{xQ}, \qquad L_x^{(0)} = QL_{xP}, \tag{46}$$

for non-singular linear transformations P and Q.

We observe that (46) implies that $R_x = P^{-1}R_{xQ^{-1}}^{(0)}$, $L_x = Q^{-1}L_{xP^{-1}}$ and thus that if $\mathfrak{A}_0$ is a principal isotope of $\mathfrak{A}$ then $\mathfrak{A}$ is a principal isotope of $\mathfrak{A}_0$. If also $R_x^{(1)} = UR_{xV}^{(0)}$ we have $R_x^{(1)} = (UP)R_{xVQ}$, $L_x^{(1)} = VL_{xU}^{(0)} = VQL_{xUP}$. Finally $\mathfrak{A}$ is a principal isotope of itself with P and Q the identity. Thus we shall again say that $\mathfrak{A}$ and $\mathfrak{A}_0$ are principal isotopes as well as that $\mathfrak{A}_0$ is a principal isotope of $\mathfrak{A}$.

Let us note in closing this section that every automorphism of an algebra $\mathfrak{A}$ is an equivalence H of $\mathfrak{A}$ and itself. Then $(a, x) = a \cdot x$ and $R_x^{(0)} = R_x$, $L_x^{(0)} = L_x$. We state this result as

THEOREM 6. *A linear transformation H on an algebra $\mathfrak{A}$ defines an automorphism of $\mathfrak{A}$ if and only if H is non-singular and such that either* (*and hence both*) *of the following conditions holds*

$$R_{xH} = H^{-1}R_xH, \qquad L_{xH} = H^{-1}L_xH \qquad (x \text{ in } \mathfrak{A}). \tag{47}$$

12. Isotopes with a unity quantity

If $\mathfrak{A}$ has a unity quantity e and f is any non-zero quantity of $\mathfrak{A}$ there exists a non-singular linear transformation H such that $e = fH$. Then (43) and (44) imply that $R_f^{(0)} = HR_eH^{-1} = L_f^{(0)} = HL_eH^{-1} = I$ since $R_e = L_e = I$. It follows that $\mathfrak{A}$ is equivalent to an algebra $\mathfrak{A}_0$ with f as unity quantity. However we seek to discover what principal isotopes of $\mathfrak{A}$ have f as unity quantity. We shall obtain the answer to this question in

THEOREM 7. *Let g range over all left non-singular quantities of $\mathfrak{A}$, h range over all right non-singular quantities of $\mathfrak{A}$, so that the non-singular linear transformations*

$$P = (R_h)^{-1}, \qquad Q = (L_g)^{-1} \tag{48}$$

exist for each g and h. Then the principal isotope of $\mathfrak{A}$ defined by (46), (48) *has $f = g \cdot h$ as a unity quantity. Conversely every isotope of $\mathfrak{A}$ with a unity quantity f is equivalent to a principal isotope determined as in* (48), (46) *for $f = g \cdot h$.*

For if $f = g \cdot h$ we have $f = gR_h = hL_g$ and $g = fP$, $h = fQ$ where P and Q are defined in (48). Let $\mathfrak{A}_0$ be the principal isotope of $\mathfrak{A}$ defined by (46) for this P and Q and put $x = f$. Then

$$R_f^{(0)} = PR_h = L_f^{(0)} = QL_g = I,$$

f is the unity quantity of $\mathfrak{A}_0$. Conversely let (46) define an isotope of $\mathfrak{A}$ with f as its unity quantity so that if we define $g = fP$, $h = fQ$ we have $R_f^{(0)} = I = PR_h$, $L_f^{(0)} = I = QL_g$. But then h is right non-singular, g is left non-singular, (54) holds and $f = gP^{-1} = gR_h = g \cdot h$ as desired.

We now prove

THEOREM 8. *Let $\mathfrak{A}$ and $\mathfrak{A}_0$ be principal isotopes and let each of these algebras have a unity quantity. Then the corresponding transformation algebras $T(\mathfrak{A})$ and $T(\mathfrak{A}_0)$ are the same.*

For we have (48) and hence have P and Q in $T(\mathfrak{A})$, $R_x^{(0)}$ and $L_x^{(0)}$ in $T(\mathfrak{A})$, $T(\mathfrak{A}_0)$ is contained in $T(\mathfrak{A})$. The converse follows by symmetry.

13. Ideals in isotopes

The mapping $x \to R_x$ of $\mathfrak{L}$ on $R(\mathfrak{A})$ is one-to-one if and only if $x \to R_x^{(0)} = PR_{xQ}C$ is one-to-one. Hence $\mathfrak{A}$ contains no absolute right divisors of zero if and only if every isotope of $\mathfrak{A}$ has this property. In particular $\mathfrak{A}$ is a zero algebra if and only if every isotope of $\mathfrak{A}$ is a zero algebra. We combine this result with those of Theorems 1, 5, 8 to obtain

THEOREM 9. *Let $\mathfrak{A}$ and $\mathfrak{A}_0$ be isotopic algebras each possessing a unity quantity. Then $\mathfrak{A}$ is simple if and only if $\mathfrak{A}_0$ is simple.*

We also have the stronger result

THEOREM 10. *Let $\mathfrak{A}$ and $\mathfrak{A}_0$ be principal isotopes and let each have a unity quantity. Then a linear subspace of $\mathfrak{L}$ is an ideal of $\mathfrak{A}$ if and only if it is an ideal of $\mathfrak{A}_0$.*

This follows from Lemma 4 and from Theorem 8. That it is desirable whenever possible to restrict our attention to algebras with a unity quantity is strongly indicated by the remarkable

THEOREM 11. *Let there exist a polynomial $f(x)$ of degree n over $\mathfrak{F}$ which is irreducible in $\mathfrak{F}$. Then every algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$ and with a unity quantity e has a principal isotope which is simple and indeed has neither left nor right ideals.*

For by Lemma 6 if we take the linear transformation P of $(\mathfrak{F})_n$ such that $f(P) = 0$ the only idempotents E such that $EP = EPE$ are $0, I$. Define $\mathfrak{A}_0$ by (46) for this P and $Q = P$. Then $R_{eP^{-1}}^{(0)} = PR_e = P$, $L_{eP^{-1}}^{(0)} = QL_e = P$ and $ER(\mathfrak{A}_0) = ER(\mathfrak{A}_0)E$ is not possible unless $EP = EPE$, $EL(\mathfrak{A}_0) = EL(\mathfrak{A}_0)E$ is not possible unless $EP = EPE$. Hence in either case $E = 0, I$, the only right and left ideals of $\mathfrak{A}_0$ are zero and $\mathfrak{A}_0$.

14. Associative algebras

It is well known[9] that if $\mathfrak{A}$ is an associative algebra with a unity quantity the space $R(\mathfrak{A})$ is an algebra and the mapping $x \to R_x$ defines an equivalence of $\mathfrak{A}$ and $R(\mathfrak{A})$. Moreover $L(\mathfrak{A})$ is also an algebra and $x \to L_x$ defines a reciprocal simple isomorphism of $\mathfrak{A}$ and $L(\mathfrak{A})$. However it is possible for $R(\mathfrak{A})$ to be an algebra without $\mathfrak{A}$ being associative. We shall give an illustration of such an occurrence shortly.

Let us now observe the known criterion for associativity which we state as

LEMMA 11. *Let $R(\mathfrak{A})$ and $L(\mathfrak{A})$ be the right and left multiplication spaces*

[9] Cf. the reference in footnote 7.

respectively of an algebra $\mathfrak{A}$. Then $\mathfrak{A}$ is associative if and only if $RL = LR$ for every R of $R(\mathfrak{A})$ and L of $L(\mathfrak{A})$.

The proof of this criterion is rather immediate. We write $(x \cdot a) \cdot y = x \cdot (a \cdot y)$ for every a, x, y of $\mathfrak{A}$ and see that this equation is equivalent to

$$(aL_x)R_y = x(aR_y) = aR_yL_x . \tag{49}$$

Thus $L_xR_y = R_yL_x$ as desired.

We now derive the important

THEOREM 12. *An algebra $\mathfrak{A}$ with a unity quantity is associative if and only if every isotope with a unity quantity of $\mathfrak{A}$ is associative and equivalent to $\mathfrak{A}$.*

For if $\mathfrak{A}$ is associative $R_xR_y = R_{xy}$, $L_xL_y = L_{yx}$ for every x and y of $\mathfrak{A}$. A quantity x of $\mathfrak{A}$ is right non-singular if and only if it has an inverse in $\mathfrak{A}$ and then x is also left non-singular. But then

$$R_{x^{-1}} = (R_x)^{-1}, \qquad L_{x^{-1}} = (L_x)^{-1}. \tag{50}$$

Let now $\mathfrak{A}_0$ be a principal isotope of $\mathfrak{A}$ and assume that $\mathfrak{A}_0$ has a unity quantity so that (48) holds. Then $P = R_{h^{-1}}$, $Q = L_{g^{-1}}$ and we have $xQ = xL_{g^{-1}} = g^{-1} \cdot x$, $PR_{xQ} = R_{h^{-1}}R_{xQ} = R_{h^{-1} \cdot g^{-1} \cdot x}$. However $f^{-1} = (g \cdot h)^{-1} = h^{-1} \cdot g^{-1}$ and we have proved that

$$R_x^{(0)} = R_{f^{-1}x} . \tag{51}$$

Similarly $L_x^{(0)} = L_{g^{-1}}L_{xP} = L_{g^{-1}}L_{x \cdot h^{-1}} = L_{x \cdot h^{-1} \cdot g^{-1}}$,

$$L_x^{(0)} = L_{x \cdot f^{-1}}. \tag{52}$$

It follows that $R(\mathfrak{A}_0) = R(\mathfrak{A})$, $L(\mathfrak{A}_0) = L(\mathfrak{A})$, $R_x^{(0)}L_y^{(0)} = L_y^{(0)}R_x^{(0)}$ for every x and y of $\mathfrak{L}$. By Lemma 11 the algebra $\mathfrak{A}_0$ is associative. Since it has a unity quantity it is equivalent to the algebra $R(\mathfrak{A}_0) = R(\mathfrak{A})$ which is equivalent to $\mathfrak{A}$, $\mathfrak{A}$ and $\mathfrak{A}_0$ are equivalent.

Observe that if $H = R_{f^{-1}}$ then $xH = x \cdot f^{-1}$, $HR_{xH}H^{-1} = R_{f^{-1}}R_{x \cdot f^{-1}}R_f = R_{f^{-1} \cdot x}$. Hence

$$R_x^{(0)} = HR_{xH}H^{-1} \tag{53}$$

and the principal isotopy of $\mathfrak{A}$ and $\mathfrak{A}_0$ which we are studying is induced by the linear mapping

$$x \to xH = xf^{-1}$$

of $\mathfrak{A}_0$ on $\mathfrak{A}$. This map is an equivalence of $\mathfrak{A}_0$ and $\mathfrak{A}$ obtainable as the product of the equivalence $x \to R_x^{(0)}$ of $\mathfrak{A}_0$ on $R(\mathfrak{A}_0) = R(\mathfrak{A})$, the automorphism $R_x^{(0)} = HR_{xH}H^{-1} \to R_{xH}$ of $R(\mathfrak{A})$ and the equivalence $R_{xH} \to xH$ of $R(\mathfrak{A})$ on $\mathfrak{A}$. Observe also that the only principal isotopy of an associative algebra with a unity quantity e which carries e into the unity quantity of the isotope is that given by $R_x^{(0)} = R_x$. For (51) holds with $f = e, f^{-1} \cdot x = e \cdot x = x$.

It is natural to consider at this point whether the property that $R(\mathfrak{A})$ is an algebra implies that $\mathfrak{A}$ is an associative algebra. This is partially true in view of

THEOREM 13. *An algebra $\mathfrak{A}$ with a left unity quantity e is associative if and only if $R(\mathfrak{A})$ is an algebra. Moreover the mapping $x \to R_x$ then is an equivalence of $\mathfrak{A}$ and $R(\mathfrak{A})$.*

For we have $e \cdot x = eR_x = x$, the mapping $R_x \to eR_x = x$ is a one-to-one linear mapping of $\mathfrak{A}$ on $R(\mathfrak{A})$. Now $(e \cdot x) \cdot y = x \cdot y = eR_xR_y = e \cdot (x \cdot y) = eR_{x \cdot y}$ and $R_{x \cdot y}$ is in $R(\mathfrak{A})$. It follows that $R_{x \cdot y} = R_xR_y$, $\mathfrak{A}$ is equivalent to $R(\mathfrak{A})$ and is associative. The converse has already been mentioned.

We may also prove the simple generalization

THEOREM 14. *Let $\mathfrak{A}$ and $R(\mathfrak{A})$ be algebras and let there be a left non-singular quantity f in $\mathfrak{A}$. Then $\mathfrak{A}$ has an associative principal isotope $\mathfrak{A}_0$ which is equivalent to $R(\mathfrak{A})$ and has f as left unity quantity.*

For we define $\mathfrak{A}_0$ by $R_x^{(0)} = R_{xQ}$, $Q = (L_f)^{-1}$. Then $L_x^{(0)} = QL_x = (L_f)^{-1}L_x$ and hence $L_f^{(0)} = I$, $(f, x) = xL_f^{(0)} = x$, $\mathfrak{A}_0$ has f as its left unity quantity. But $R(\mathfrak{A}) = R(\mathfrak{A}_0)$ and our result follows from Theorem 13.

It remains to consider the general question as to the existence of non-associative algebras $\mathfrak{A}$ such that $R(\mathfrak{A})$ is an algebra. Such algebras do exist and we prove this as an immediate consequence of

THEOREM 15. *Let $\mathfrak{A}$ be an associative algebra of order $n > 1$ over an infinite field $\mathfrak{F}$ and let $\mathfrak{A}$ have a unity quantity e so that $R(\mathfrak{A})$ is an algebra. Then there exists a non-associative isotope $\mathfrak{A}_0$ of $\mathfrak{A}$ with $R(\mathfrak{A}_0) = R(\mathfrak{A})$.*

For it is known[9] that $L(\mathfrak{A})$ is the $(\mathfrak{F})_n$-centralizer of $R(\mathfrak{A})$, $L(\mathfrak{A})$ is a proper subalgebra of $(\mathfrak{F})_n$. Then there exists a linear transformation U not in $L(\mathfrak{A})$, $UR_a - R_aU \neq 0$ for some a in $\mathfrak{A}$. Every linear transformation has the form

$$U = \sum_{i=1}^{n^2} \xi_i S_i \qquad (\xi_i \text{ in } \mathfrak{F})$$

for a basis S_i of $(\mathfrak{F})_n$, and $UR_a - R_aU \neq 0$ implies that there exist η_i in $\mathfrak{F}$ such that $Q = \sum \eta_i S_i$ is non-singular, $QR_a \neq R_aQ$. We define $\mathfrak{A}_0$ by $R_x^{(0)} = R_{xQ}$ and have $R(\mathfrak{A}_0) = R(\mathfrak{A})$, $L_x^{(0)} = QL_x$, $L_e^{(0)} = Q$ is not commutative with $R_{aQ^{-1}}^{(0)}$, $\mathfrak{A}_0$ is not associative.

It is important to observe that *there exist associative algebras* (without unity quantities) *which are isotopic but not equivalent.* For example consider the nilpotent algebra $\mathfrak{A}$ with basis e_1, e_2, e_3 such that $e_1 \cdot e_2 = -e_2 \cdot e_1 = e_3$ and all other products are zero. Then we write $a = (\alpha_1, \alpha_2, \alpha_3)$ for $a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$ and have $a \cdot x = a \cdot (\xi_1, \xi_2, \xi_3) = (0, 0, \alpha_1\xi_2 - \alpha_2\xi_1) = a \cdot \Gamma_x$, $x \cdot a = (0, 0, \xi_1\alpha_2 - \xi_2\alpha_1) = a\Delta_x$, where

$$\Gamma_x = -\Delta_x = \begin{pmatrix} 0 & 0 & \xi_2 \\ 0 & 0 & -\xi_1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{54}$$

This algebra is associative since

$$(55) \qquad \Gamma_x \Delta_y = \begin{pmatrix} 0 & 0 & \xi_2 \\ 0 & 0 & -\xi_1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & -\eta_2 \\ 0 & 0 & \eta_1 \\ 0 & 0 & 0 \end{pmatrix} = 0 = \Delta_y \Gamma_x .$$

We let

$$(56) \qquad \Lambda = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and define

$$(57) \qquad \Gamma_x^{(0)} = \Lambda \Gamma_x = \begin{pmatrix} 0 & 0 & \xi_1 \\ 0 & 0 & \xi_2 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \Delta_x^{(0)} = \Delta_{x\Lambda} = \begin{pmatrix} 0 & 0 & \xi_1 \\ 0 & 0 & \xi_2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $\Gamma_x^{(0)} \Delta_x^{(0)} = \Delta_x^{(0)} \Gamma_x^{(0)} = 0$ as before. But we have then defined an isotope $\mathfrak{A}_0$ of $\mathfrak{A}$ with $(e_1, e_1) = e_1 \Gamma e_1^{(0)} = e_3$. It is not equivalent to $\mathfrak{A}$ since in $\mathfrak{A}$ the square of every a is $a\Gamma_a = (0, 0, \alpha_1\alpha_2 - \alpha_2\alpha_1) = 0$.

15. Isotopes with a prescribed unity quantity

Let $\mathfrak{N}$ be a linear subspace of order n over $\mathfrak{F}$ of $(\mathfrak{F})_n$ and I be in $\mathfrak{N}$. Then we have seen that if f is any non-zero vector of $\mathfrak{L}$ and the linear mapping

$$(58) \qquad R \rightarrow fR$$

of $\mathfrak{N}$ on $\mathfrak{L}$ is one-to-one the algebra $\mathfrak{A}$ with $R(\mathfrak{A}) = \mathfrak{N}$ and which is defined by

$$(59) \qquad fS \cdot fR = f \cdot SR \qquad (S \text{ in } (\mathfrak{F})_n, R \text{ in } \mathfrak{N})$$

has f as left unity quantity, $L_f = I$. But also if $fR = x$ we have $R = R_x$. Since $\mathfrak{N}$ contains I we have $fI = f$, $I = R_f$, f is the unity quantity of $\mathfrak{A}$.

Conversely if $\mathfrak{A}$ has f as its unity quantity we have $f \cdot x = fR_x = x$ and the linear mapping $R_x \rightarrow fR_x$ is one-to-one and is such that $I = R_f$ is in $\mathfrak{N} = R(\mathfrak{A})$.

Let us now assume that $\mathfrak{A}$ is a prescribed algebra with a unity quantity e so that $R(\mathfrak{A})$ contains I. We now let P and C be non-singular linear transformations and G in $R(\mathfrak{A})$ have the property that

$$(60) \qquad PGC = I.$$

Then $\mathfrak{N} = PR(\mathfrak{A})C$ contains I and if we let f be any vector such that the linear mapping

$$N \rightarrow fN$$

of $\mathfrak{N}$ on $\mathfrak{L}$ is one-to-one we define an algebra $\mathfrak{A}_0$ with $R(\mathfrak{A}_0) = \mathfrak{N}$ by $fS \cdot fN = fSN$ for every S in $(\mathfrak{F})_n$ and N in $\mathfrak{N}$ and have seen that f is its unity quantity. It is then desirable to have

THEOREM 16. *The algebra $\mathfrak{A}_0$ with unity quantity f defined above is an isotope of $\mathfrak{A}$ with f as unity quantity.*

For $fN = x$ implies that $N = R_x^{(0)} = PR_zC$. Write $g = fP$ and have $x = gR_zC = (g \cdot z)C = zL_gC$. If L_g is singular so is L_gC, that is the mapping of N on $fN = x$ is singular. It follows that $N \to fN$ is non-singular if and only if $g = fP$ is a left non-singular quantity of $\mathfrak{A}$. We put $Q = C^{-1}(L_g)^{-1}$ and have $z = xQ$, $R_x^{(0)} = PR_{xQ}C$ as desired.

16. Commutative isotopes

An algebra $\mathfrak{A}$ is commutative if and only if $a \cdot x = aR_x = x \cdot a = aL_x$, that is,

$$R_x = L_x \tag{61}$$

for every x of $\mathfrak{A}$. Let $\mathfrak{A}_0$ be a principal isotope of an algebra $\mathfrak{A}$ (which may or may not be commutative) so that $\mathfrak{A}_0$ is commutative if and only if $PR_{xQ} = QL_{xP}$ for every x. Put $y = xP$ and have $xQ = yP^{-1}Q = yS$ where $S = P^{-1}Q$. Thus $\mathfrak{A}_0$ is commutative if and only if

$$R_{yS} = SL_y \tag{62}$$

for every y of $\mathfrak{A}$.

Suppose now that $\mathfrak{A}$ has a left unity quantity e, that is, $e \cdot x = x$ for every x of $\mathfrak{A}$, $L_e = I$. Then (62) implies that $S = R_f$. Thus f is a right non-singular quantity of $\mathfrak{A}$. Moreover $yS = yR_f = y \cdot f$ and (62) is equivalent to $x \cdot (yS) = xR_fL_y = y \cdot (x \cdot f)$, that is, to

$$x \cdot (y \cdot f) = y \cdot (x \cdot f) \tag{63}$$

for every x and y of $\mathfrak{A}$.

Conversely let P be any non-singular linear transformation, f be any right non-singular quantity of $\mathfrak{A}$ and put $Q = PR_f$ so that $S = R_f = P^{-1}Q$ and (63) implies (62). We have proved

THEOREM 17. *Let $\mathfrak{A}$ be an algebra with a left unity quantity, f range over all right non-singular quantities of $\mathfrak{A}$ such that* (63) *holds for every x and y of $\mathfrak{A}$. Then the principal isotopes of $\mathfrak{A}$ defined by $R_x^{(0)} = PR_{xPR_f}$, for P any non-singular quantity of $(\mathfrak{F})_n$, are commutative algebras to one of which every commutative isotope of $\mathfrak{A}$ is equivalent.*

17. Division algebras

An algebra $\mathfrak{A}$ is called a *division algebra* if it has no (right and hence no left) divisors of zero. Then every non-zero quantity of $\mathfrak{A}$ is both left and right non-singular and we apply Theorem 7 with $g = h \neq 0$ to obtain as an immediate consequence.[10]

[10] This result was also obtained by Steenrod. His proof was necessarily more complicated as he did not have Theorems 5 and 7.

THEOREM 18. *Every division algebra is isotopic to a division algebra with a unity quantity.*

Non-associative division algebras do not have many of the properties of associative division algebras. In particular the right minimum function of a quantity of a division algebra may be reducible. Let us note now that a division algebra $\mathfrak{A}$ has no right ideals other than $\mathfrak{A}$ or zero. For otherwise we would have $b \cdot x$ in a right ideal $\mathfrak{B}$ for every b of $\mathfrak{B}$ and x of $\mathfrak{A}$ whereas $b \cdot x = a$ has the solution $x = a(L_b)^{-1}$ for every a of $\mathfrak{A}$. Similarly $\mathfrak{A}$ has no left ideals other than $\mathfrak{A}$ or zero.

If $\mathfrak{B}$ is a division subalgebra of an algebra $\mathfrak{A}$ and the unity quantity of $\mathfrak{A}$ is in $\mathfrak{B}$ we may prove that if u is any quantity in $\mathfrak{A}$ and not in $\mathfrak{B}$ the linear spaces $\mathfrak{B}$, $u\mathfrak{B}$ are supplementary in their sum. For otherwise $u \cdot b = b_0$ for non-zero b and b_0 in $\mathfrak{B}$, $u = b_0(R_b)^{-1}$. But the equation $xb = b_0$ has the solution $x = b_0(R_b)^{-1}$ in the division algebra $\mathfrak{B}$ and u is in $\mathfrak{B}$, a contradiction.

The process above is used for associative algebras to prove the theorem that the order of $\mathfrak{B}$ divides the order of $\mathfrak{A}$ under the hypothesis just stated. The usual proof of this result requires that if $u_1\mathfrak{B} + \cdots + u_r\mathfrak{B} = \mathfrak{B}_r$ is a supplementary sum of linear spaces not containing u then so is the sum of $\mathfrak{B}_r$ and $u\mathfrak{B}$. Thus in particular we need to show that if u is not in $v\mathfrak{B}$ no non-zero quantity of $u\mathfrak{B}$ is in $v\mathfrak{B}$. But $u \cdot b = v \cdot b_0$ then $u = (v \cdot b_0)R_b^{-1} = vR_{b_0}(R_b)^{-1}$. However we cannot conclude that this latter quantity is in $v\mathfrak{B}$ and thus our proof breaks down. We leave the question as to the validity of this theorem as an unsolved problem.

If $\mathfrak{A}$ is a division algebra every R_x defined for $x \neq 0$ is non-singular and $fR_x = 0$ if and only if $f = 0$. Thus the mapping $R \to fR$ of $R(\mathfrak{A})$ on $\mathfrak{A}$ is non-singular for every $f \neq 0$. Moreover so is the mapping $PRC \to fPRC$ for every non-singular P and C. By Theorem 16 we have

THEOREM 19. *Let f be any non-zero quantity of a division algebra $\mathfrak{A}$ and P and Q be any non-singular linear transformations such that $PRQ = I$ for some R of $R(\mathfrak{A})$. Then the algebra $\mathfrak{A}_0$ defined by $(a, fS) = aS$ for every S of $PR(\mathfrak{A})Q$ is an isotope of $\mathfrak{A}$ with f as unity quantity.*

If $\phi(\lambda)$ is the right minimum function of a quantity b in a division algebra $\mathfrak{A}$ with a unity quantity e and $\phi(\lambda)$ is reducible and of degree $t > 1$ it cannot have a linear factor. For otherwise $\phi(\lambda) = \psi(\lambda)[\lambda - \alpha]$ and $\phi_R(b) = e\phi(R_b) = [e\psi(R_b)](R_b - \alpha I) = \psi_R(b) \cdot (b - \alpha e) = 0$ which is impossible since $\psi_R(b) \neq 0$, $b - \alpha e \neq 0$. However it is possible that $\phi(\lambda) = \psi(\lambda)(\lambda^2 + \alpha\lambda + \beta)$ since then $\phi_R(b) = [e\psi(R_b)](R_b^2 - \alpha R_b + \beta I) \neq [\psi(b)] \cdot (b^2 - \alpha b + \beta e)$ since in general $R_b^2 \neq R_{b^2}$. It then becomes of interest to ask whether or not *any* quantity of a division algebra has irreducible right minimum function. It is not easy to answer this but we may prove instead

THEOREM 20. *Let $\mathfrak{A}$ be a division algebra with a unity quantity e over $\mathfrak{F}$ and let b in $\mathfrak{A}$ be not in $\mathfrak{F}$. Then there exists an isotope $\mathfrak{A}_0$ of $\mathfrak{A}$ such that $\mathfrak{A}_0$ has a unity quantity, $R(\mathfrak{A}_0) = R(\mathfrak{A})$, the right minimum function of b in $\mathfrak{A}_0$ is irreducible.*

For it suffices to assume that the right minimum function $\phi(\lambda)$ of b is reducible,

$\phi(\lambda) = \pi(\lambda)\psi(\lambda)$ where $\psi(\lambda)$ is irreducible and has degree $t > 1$. Then $\phi_R(b) = e\phi(R_b) = e\pi(R_b)\cdot\psi(R_b) = f\psi(R_b) = 0$ where $f = e\pi(R_b) = \pi_R(b) \neq 0$. We pass to the isotope $\mathfrak{A}_0$ defined as in Theorem 19 for $P = Q = I$, $R(\mathfrak{A}_0) = R(\mathfrak{A})$ and have f as the unity quantity of $\mathfrak{A}_0$. Then $f\psi(R_b) = \psi_R(b) = 0$ in $\mathfrak{A}_0$. By Theorem 2 the right minimum function of b divides $\psi(\lambda)$ and must coincide with this irreducible polynomial.

The result just obtained implies that every division algebra of order $n > 1$ over the field $\mathfrak{R}'$ of all real numbers is isotopic to a division algebra with a unity quantity e and containing a quantity b such that $b^2 = -e$. Moreover it is clear that every division algebra $\mathfrak{A}$ of order $n > 2$ over $\mathfrak{R}'$ is central simple. For otherwise we could write $\mathfrak{A}$ as a division algebra over its center $\mathfrak{C} \neq \mathfrak{R}'$, $\mathfrak{C}$ must be $\mathfrak{R}'(i)$ for $i^2 = -1$, $\mathfrak{A}$ over $\mathfrak{C}$ has an isotope $\mathfrak{A}_0$ over $\mathfrak{C}$ such that b in $\mathfrak{A}_0$ has $\lambda^2 + 1$ as (right) minimum function. But $\lambda^2 + 1 = (\lambda + i)(\lambda - i)$ in $\mathfrak{C}$ contrary to the proof above.

18. Subalgebras of isotopes

The problem of finding in a division algebra a quantity whose right minimum function is irreducible is an instance of the problem of determining whether an algebra $\mathfrak{A}$ has a certain type of subalgebra. In particular we may ask whether or not a given algebra $\mathfrak{A}$ has any proper subalgebras. A criterion that this be the case was given in Lemma 2 and we wish now to propose the question as to whether a principal isotope of $\mathfrak{A}$ has subalgebras of the same order as those of $\mathfrak{A}$. By Lemma 4 we have $\mathfrak{B} = \mathfrak{A}E$ is a subalgebra of $\mathfrak{A}$ whose order is the rank of $E \neq 0$ if and only if $ER_y = ER_yE$, $EL_y = EL_yE$ for $y = xE$ and every x of $\mathfrak{A}$. Now $R_x^{(0)} = PR_{xQ}$, $L_x^{(0)} = QL_{xP}$ and $\mathfrak{A}E_0$ is a subalgebra of $\mathfrak{A}_0$ if and only if $E_0R_z^{(0)} = E_0R_z^{(0)}E_0$, $E_0L_z^{(0)} = E_0L_z^{(0)}E_0$ for every x of $\mathfrak{L}$ where $z = xE_0$.

The problem just proposed does not appear to have a simple solution for arbitrary algebras. However we should observe that if $\mathfrak{A}_0$ has a unity quantity then $P = (R_h)^{-1}$, $Q = (L_h)^{-1}$ and if g and h are in $\mathfrak{B}$ the linear space $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}_0$ as well as of $\mathfrak{A}$. For P and Q are in $T(\mathfrak{B}, \mathfrak{A}) = ET(\mathfrak{B}, \mathfrak{A})E$ and $ER_y^{(0)} = EPR_{yQ} = EPER_{yQ} = EPER_{yQ}E = ER_y^{(0)}$ since $yQ = xEQ = xEQE = yQE$ is in $\mathfrak{B}$. Similarly $EL_y^{(0)} = EL_y^{(0)}E$. We shall not study the general question further except to note that if E_0 has the same rank as E it has the form $H^{-1}EH$ and it may be seen that $\mathfrak{B}_0 = \mathfrak{A}E_0$ is a subalgebra of $\mathfrak{A}_0$ if and only if $\mathfrak{B}$ is a subalgebra of the isotope $\mathfrak{A}_1$ defined by $R_x^{(1)} = HR_{xH}^{(0)}H^{-1}$ and equivalent to $\mathfrak{A}_0$.

19. Special properties

An algebra $\mathfrak{A}$ is said to be alternative if $(a\cdot x)\cdot x = a\cdot(x\cdot x)$, $x\cdot(x\cdot a) = (x\cdot x)\cdot a$ for every x and a of $\mathfrak{A}$. Then $\mathfrak{A}$ *is alternative if and only if*

$$R_{x^2} = (R_x)^2, \qquad L_{x^2} = (L_x)^2 \tag{64}$$

for every x *of* $\mathfrak{A}$.

It follows from (64) that $x\cdot x^2 = x(R_x)^2 = (xR_x)R_x = x^2\cdot x$. Suppose then

that $R_{x^k} = (R_x)^k$ for all right powers $k = 1, 2, \cdots, t$ and that $x \cdot x^k = x^k \cdot x$ for $k = 1, \cdots, t$. Then we put $y = x + x^t$ and have $y^2 = x^2 + (x^t)^2 + x \cdot x^t + x^t \cdot x = x^2 + (x^t)^2 + 2x^{t+1}$. But $R_y = R_x + R_{x^t} = R_x + (R_x)^t$, $(R_y)^2 = (R_x)^2 + 2(R_x)^{t+1} + (R_x)^{2t} = R_{x^2} + R_{x^t \cdot x^t} + 2(R_x)^{t+1}$. It follows that $2R_{x^{t+1}} = 2(R_x)^{t+1}$ and that $(R_x)^{t+1} = R_{x^{t+1}}$ if the characteristic of $\mathfrak{F}$ is not two. But then $x \cdot x^{t+1} = x(R_x)^{t+1} = (xR_{x^t})R_x = x^{t+1} \cdot x$. This completes our induction and proves that $R_{x^k} = (R_x)^k$ for every k.

We see that consequently $x^k \cdot x^t = xR_x^{s+t-1} = x^{s+t}$ so that all powers of x are right powers, $(x^s \cdot x^t) \cdot x^k = x^{s+t+k} = x^s \cdot (x^t \cdot x^k)$. It follows that the algebra $\mathfrak{F}[x]$ of all right polynomials $\phi_R(x)$ is the associative algebra of all polynomials $\phi(x) = \phi_R(x) = \phi_L(x)$ and $R_{\phi(x)} = \phi(R_x)$, $L_{\phi(x)} = \phi(L_x)$.

We now propose the problem of determining the principal isotopes of an algebra $\mathfrak{A}$ which are alternative. This occurs if and only if $(R_z^{(0)})^2 = R_z^{(0)}$, $(L_z^{(0)})^2 = L_z^{(0)}$ where $z = xR_x^{(0)} = xL_x^{(0)}$. But then we must have

$$R_{xQ}PR_{xQ} = R_{zQ}\,, \qquad L_{xP}QL_{xP} = L_{zP}\,.$$

Replace xQ by x and thus $zQ = xQL_{xP}Q$ by $(xQ^{-1}P \cdot x)Q$ and similarly replace xP by x and thus $zP = xPR_{xQ}P$ by $(xP^{-1}Q \cdot x)P$. Then we see that $\mathfrak{A}_0$ is alternative if and only if

$$R_xPR_x = R_u\,, \qquad L_xQL_x = L_v\,,$$

for every x of $\mathfrak{A}$, where

$$u = (xQ^{-1}P \cdot x)Q, \qquad v = (xP^{-1}Q \cdot x)P,$$

and the indicated products are those in $\mathfrak{A}$. It is of particular interest, of course, to study the case where we assume also that $\mathfrak{A}$ is alternative.

An algebra $\mathfrak{A}$ is called a *Lie algebra* if $a \cdot x = -x \cdot a$, $a \cdot (x \cdot y) + y \cdot (a \cdot x) + x \cdot (y \cdot a) = 0$ for every a, x, y of $\mathfrak{A}$. Then $L_x = -R_x$, $aR_{x \cdot y} + aR_xL_y + aL_yL_x = 0$, $a[R_{xy} - (R_xR_y - R_yR_x)] = 0$, $\mathfrak{A}$ is a Lie algebra if and only if

$$L_x = -R_x\,, \qquad R_{x \cdot y} = R_xR_y - R_yR_x \qquad (x,\ y \text{ in } \mathfrak{A}). \tag{65}$$

We propose again the question as to whether a principal isotope of an algebra $\mathfrak{A}$ is a Lie algebra and see that this occurs if and only if

$$PR_{xQ} = -QL_{xP}\,, \qquad PR_{xQ}PR_{yQ} - PR_{yQ}PR_{xQ} = PR_z$$

where $z = (x, y)Q = yQL_{xP}Q$. Replace xQ by x, yQ by y and thus xP by xC for $C = Q^{-1}P$, z by $yL_{xC}Q = (xC \cdot y)Q$. Then $\mathfrak{A}_0$ is a Lie algebra if and only if

$$L_{xC} = -CR_x\,, \qquad R_xPR_y - R_yPR_x = R_{(xC \cdot y)Q}\,. \tag{66}$$

The problem of determining the principal Lie isotopes of simple Lie algebras is being studied.[11]

[11] This problem is the topic of study of a doctoral dissertation at the University of Chicago. We also wish to mention here that Mr. W. Carter in his Master's dissertation has classified all real division algebras of order four and degree two into classes of algebras

Let us conclude these remarks with some observations which will be important for the study of simple algebras. We define the *center*[12] of any algebra $\mathfrak{A}$ over $\mathfrak{F}$ to be the set $\mathfrak{Z}$ of all quantities z of $\mathfrak{A}$ such that $R_z = L_z$ is commutative with every R_x of $R(\mathfrak{A})$ and L_x of $L(\mathfrak{A})$. Then z is in $\mathfrak{Z}$ if and only if

$$z \cdot a = a \cdot z, \quad z \cdot (a \cdot x) = (z \cdot a) \cdot x, \quad a \cdot (z \cdot x) = (a \cdot z) \cdot x, \quad (a \cdot x) \cdot z = a \cdot (x \cdot z)$$

for every a and x of $\mathfrak{A}$. It is easily shown that $\mathfrak{Z}$ is zero or an associative subalgebra of $\mathfrak{A}$. Moreover if $\mathfrak{A}$ has a unity quantity e the set $\mathfrak{Y} = e\mathfrak{F}$ of all αe for α in $\mathfrak{F}$ is a subalgebra of order one over $\mathfrak{F}$ of $\mathfrak{Z}$, $e\mathfrak{F}$ is equivalent to $\mathfrak{F}$ and is a field.[13]

Let us define a new operation of scalar product on $\mathfrak{A}\mathfrak{Y}$ to $\mathfrak{A}$ by writing $(a, y) = a \cdot y = a \cdot \alpha e = a\alpha$ for every a of $\mathfrak{A}$ and α of $\mathfrak{F}$. Then $\mathfrak{A}$ is an algebra over $\mathfrak{Y}$ with respect to this operation. It is clear that this is a change in our representation of $\mathfrak{A}$ as a linear space over a field and is not a change in $\mathfrak{A}$.

If $\mathfrak{A}$ is a simple algebra of order n over $\mathfrak{F}$ we have seen that $\mathfrak{A}$ is a central simple algebra of order s over its transformation center $\mathfrak{C}$, $\mathfrak{C}$ is a field of degree t over $\mathfrak{F}$, $n = st$. Let $\mathfrak{A}$ have e as its unity quantity so that, as above, $e\mathfrak{C}$ is a subalgebra over $\mathfrak{C}$ of $\mathfrak{A}$ and is equivalent to $\mathfrak{C}$. But then $e\mathfrak{C}$ is a field of degree t over $e\mathfrak{F}$, $e\mathfrak{C}$ is a subalgebra of $\mathfrak{A}$ of order t over $\mathfrak{F}$. Moreover it is easy to verify that $e\mathfrak{C}$ is contained in the center $\mathfrak{Z}$ of $\mathfrak{A}$. But the quantities of $\mathfrak{Z}$ are quantities z such that R_z is in $\mathfrak{C}$, $e \cdot z = eR_z = z$ is in $e\mathfrak{C}$. This proves that $e\mathfrak{C}$ is the center of every simple algebra $\mathfrak{A}$ with e as unity quantity and $\mathfrak{C}$ as its transformation center. If we express $\mathfrak{A}$ as an algebra over $\mathfrak{C}$ it is central simple and consequently we may express $\mathfrak{A}$ as a central simple algebra over the associative subalgebra of $\mathfrak{A}$ which is its center $\mathfrak{Z}$.

It is desirable to note that if $\mathfrak{A}$ and $\mathfrak{A}_0$ are principal isotopes we have $T(\mathfrak{A}) = T(\mathfrak{A}_0)$ and hence $\mathfrak{C} = \mathfrak{C}(\mathfrak{A}) = \mathfrak{C}(\mathfrak{A}_0)$, $\mathfrak{A}$ and $\mathfrak{A}_0$ have the same transformation center. If e and e_0 are corresponding unity quantities we have $e\mathfrak{C}$ equivalent over $\mathfrak{F}$ to $e_0\mathfrak{C}$ and thus we have shown that *isotopic simple algebras with unity quantities have equivalent centers*. It is important to observe that, while the center of an associative simple algebra $\mathfrak{A}$ is its $\mathfrak{A}$-centralizer, this may not be the case when $\mathfrak{A}$ is not associative.

UNIVERSITY OF CHICAGO

with respect to isotopy. He has also shown that *every real division algebra of order four is isotopic to an algebra of degree four,* that is, containing a quantity whose right minimum function has degree four and is thus reducible. Moreover *there exist real division algebras of degree and order four not isotopic to algebras of degree two.*

[12] We have now used the terms *center* instead of *centrum* and *central* in place of *normal*. A change of terminology of this kind has long seemed very desirable to many algebraists.

[13] In particular $\mathfrak{A}$ may be commutative and may yet be a central simple algebra. For example the set of all r-rowed real symmetric square matrices forms a commutative central simple algebra with respect to the product operation $a \cdot x = \frac{1}{2}(ax + xa)$, ax and xa the ordinary matrix products.

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

NON-ASSOCIATIVE ALGEBRAS

II. New Simple Algebras[1]

BY A. A. ALBERT

(Received January 30, 1942)

1. Introduction

In the second part of our study of non-associative algebras we shall give an iterative construction of new simple algebras with a unity quantity. All previous constructions[2] of this type have used groups of automorphisms or anti-automorphisms and the great generality of our definition will lie precisely in that we shall be able to use instead almost[3] arbitrary multiplicative groups of non-singular linear transformation.

We shall begin our exposition with a preliminary discussion of (non-associative) separable algebras, that is algebras $\mathfrak{A}$ with a unity quantity e such that every scalar extension of $\mathfrak{A}$ is a direct sum of simple algebras. Let $\mathfrak{G}$ be any finite multiplicative group of non-singular linear transformations S on $\mathfrak{A}$ such that $eS = e$ and $\mathfrak{H}$ be any subset of $\mathfrak{G}$ containing the identity transformation. We define an *extension set* $\mathfrak{g}$ to be a set of non-singular quantities $g_{S,T}$ in $\mathfrak{A}$ for every S and T in $\mathfrak{G}$. Then we shall construct a corresponding *crossed extension* $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ which is a certain algebra having e as its unity quantity.

For every separable $\mathfrak{A}$ we shall give conditions that the crossed extensions shall be simple (or central simple) algebras. If $\mathfrak{H}$ is the identity group $\mathfrak{C}$ is simple whenever $\mathfrak{A}$ is, $\mathfrak{C}$ is central simple whenever $\mathfrak{A}$ is. These latter algebras include the so-called *Cayley algebras*. Our algebras are associative only when $\mathfrak{G} = \mathfrak{H}$ is a group of automorphisms and our definition then includes that of crossed products.[4]

The crossed products are associative central simple algebras of order r^2, and for each such algebra we may use an explicit process to give a set of corresponding central simple algebras of order r^t for any integer $t > 1$. These algebras are not associative for $t > 2$. In particular we then have generalized cyclic algebras. Another explicit construction will connect every central simple

[1] Presented to the Society February 28, 1942.

[2] Automorphisms were necessarily used in the constructions of associative algebras of L. E. Dickson for which see my *Structure of Algebras*, Chapter V, Chapter XI, pp. 182–8, and bibliographical references [141], [145]. Cayley algebras were generalized in my "Quadratic forms permitting composition" these Annals, vol. 41, pp. 161–77, to algebras of order 2^t obtained by the process given here where the group consists of the identity automorphism and an antiautomorphism of order two.

[3] The special restrictions will reduce to the property that the transformations leave the unity quantity unaltered in the most important cases.

[4] For these algebras and the Cayley algebras see the references in footnote 2.

algebra of order n with a crossed extension of it by a group $\mathfrak{G}$ equivalent to any permutation group on m letters.

We shall close our discussion with a list of fundamental unsolved problems in the theory of these new algebras.

2. Decomposition of algebras with a unity quantity

An algebra $\mathfrak{A}$ is said to be *decomposable*[5] if it is expressible as the supplementary sum $\mathfrak{A}_1 + \cdots + \mathfrak{A}_r$, of at least two subalgebras $\mathfrak{A}_i$ of $\mathfrak{A}$, such that $a_i a_j = 0$ for a_i in $\mathfrak{A}_i$, a_j in $\mathfrak{A}_j$ and all $i \neq j$. Then we say that $\mathfrak{A}$ decomposes into the *direct sum* of its *components* $\mathfrak{A}_i$ (which are ideals of $\mathfrak{A}$), we write

$$\mathfrak{A} = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_r, \tag{1}$$

and we call (1) a *decomposition* of A. If $\mathfrak{A}$ has no such decomposition we say that $\mathfrak{A}$ is *indecomposable*. It is clear that a decomposition with r components becomes one with $r + 1$ components if we replace a decomposable component $\mathfrak{A}_i = \mathfrak{B} \oplus \mathfrak{C}$ by $\mathfrak{B} \oplus \mathfrak{C}$ in (1). The lowering of orders in such decomposition implies that every $\mathfrak{A}$ has a decomposition (1) with the components indecomposable algebras. It is then natural to ask (as in the theory of associative algebras) whether or not such a decomposition is unique apart from the ordering of the components. We shall give a simple solution below for algebras with a unity quantity.

The *center* of an algebra $\mathfrak{A}$ has been defined[6] to be the set $\mathfrak{Z}$ of all quantities z of $\mathfrak{A}$ such that the commutative and associative laws, for products in $\mathfrak{A}$, hold whenever z is one of the factors. Then $\mathfrak{Z}$ is zero or an associative and commutative subalgebra of $\mathfrak{A}$. When $\mathfrak{A}$ has a unity quantity e the subalgebra $e\mathfrak{F}$ of $\mathfrak{A}$ is in its center $\mathfrak{Z}$ and we call it a *central algebra* if $\mathfrak{Z} = e\mathfrak{F}$. We may now prove

LEMMA 1. *Let $\mathfrak{A}$ be an algebra with a unity quantity e and $\mathfrak{Z}$ be the center of $\mathfrak{A}$. Then $\mathfrak{A}$ has the form* (1) *if and only if $e = e_1 + \cdots + e_r$ for pairwise orthogonal idempotents e_i in $\mathfrak{Z}$ such that $\mathfrak{A}_i = \mathfrak{A}e_i$.*

For if $\mathfrak{A}$ has the form (1) we may write every quantity of $\mathfrak{A}$ in the form $a = a_1 + \cdots + a_r$, for the a_i uniquely determined quantities of $\mathfrak{A}_i$. Then if $x = x_1 + \cdots + x_r$ we have $a \cdot x = a_1 \cdot x_1 + \cdots + a_r \cdot x_r$. Take $x = e = e_1 + \cdots + e_r$ and have $e_i \cdot e_j = 0$ for $i \neq j$, $a \cdot e = a_1 \cdot e_1 + \cdots + a_r \cdot e_r = a$ if and only if $a_i \cdot e_i = a_i$. Similarly $e_i \cdot a_i = a_i$, $\mathfrak{A}_i$ has e_i as its unity quantity, $\mathfrak{A}_i = \mathfrak{A}e_i$. Now $(a \cdot x) \cdot e_i = (a_i \cdot x_i) \cdot e_i = a_i \cdot x_i = (a \cdot e_i) \cdot x_i = a \cdot (xe_i)$ and similar other verifications imply that e_i is in $\mathfrak{Z}$. Conversely if $e = e_1 + \cdots + e_r$, for pairwise orthogonal idempotents e_i in $\mathfrak{Z}$, we have $\mathfrak{A} = \mathfrak{A}_1 + \cdots + \mathfrak{A}_r$ for $\mathfrak{A}_i = \mathfrak{A}e_i$ and we have (1).

As a consequence of this result we have

---

[5] This term seems much preferable to the term reducible which causes so much confusion if representation theory and linear algebra theory be considered together.

[6] This was defined at the end of part I of this paper. We shall use the concepts given in that part without any reference.

LEMMA 2. *The algebra $\mathfrak{A}$ of Lemma 1 has a decomposition* (1) *if and only if its center $\mathfrak{Z}$ has a corresponding decomposition*

$$\mathfrak{Z} = \mathfrak{Z}_1 + \cdots + \mathfrak{Z}_r. \tag{2}$$

Then $\mathfrak{Z}_i$ is the intersection of $\mathfrak{A}_i$ and $\mathfrak{Z}$ and is the center of $\mathfrak{A}_i$.

For e is in $\mathfrak{Z}$ and Lemma 1 implies that from (1) we have (2) with

$$\mathfrak{Z}_i = \mathfrak{Z}e_i \tag{3}$$

and conversely (2) and (3) imply (1) with $\mathfrak{A} = \mathfrak{A}e_i$. Then $\mathfrak{Z}_i$ is in both $\mathfrak{Z}$ and $\mathfrak{A}_i$, and is the intersection of $\mathfrak{Z}$ and $\mathfrak{A}_i$ by (2). If z_i is in the center of $\mathfrak{A}_i$ then $z_i \cdot a_j = a_j \cdot z_i = 0$ for a_j in $\mathfrak{A}_j$ and $j \neq i$, z_i is in $\mathfrak{Z}_i$, $\mathfrak{Z}_i$ is the center of $\mathfrak{A}_i$.

Lemma 2 clearly implies

LEMMA 3. *An algebra $\mathfrak{A}$ with a unity quantity is indecomposable if and only if its center is indecomposable.*

We then have

LEMMA 4. *The decomposition of an algebra with a unity quantity as a direct sum* (1) *of indecomposable components is unique apart from the arrangement of the components.*

This follows from Lemmas 2, 3, and the associative case of the result we are proving. This latter result is proved by the use of the following lemma which may then be used to prove Lemma 4.

LEMMA 5. *Let an algebra $\mathfrak{A}$ with a unity quantity have a decomposition* (1) *so that $\mathfrak{A}_i = \mathfrak{A}e_i$ with e_i an idempotent of $\mathfrak{A}$. Then every right, left or two-sided ideal $\mathfrak{B}$ of $\mathfrak{A}$ is the direct sum*

$$\mathfrak{B} = \mathfrak{B}_1 \oplus \cdots \oplus \mathfrak{B}_r, \tag{4}$$

where $\mathfrak{B}_i = \mathfrak{B}e_i$ is the intersection of $\mathfrak{B}$ and $\mathfrak{A}_i$, $\mathfrak{B}_i$ is correspondingly a right, left, or two-sided ideal of $\mathfrak{A}_i$.

The proof of this result involves the use of the associative law only for products $a \cdot b \cdot c$ with a factor in the center, and thus the proof which has been given in the associative case is valid without change. We shall not repeat it here.

3. Absolute indecomposability

If $\mathfrak{Z}$ is the center of an algebra $\mathfrak{A}$ over $\mathfrak{F}$ and $\mathfrak{K}$ is any scalar extension of $\mathfrak{F}$ the center of $\mathfrak{A}_\mathfrak{K}$ is $\mathfrak{Z}_\mathfrak{K}$. For it is clear that $\mathfrak{Z}_\mathfrak{K}$ is contained in the center $\mathfrak{Z}_0$ of $\mathfrak{A}_\mathfrak{K}$. Let then z_0 be in $\mathfrak{Z}_0$ so that we may write $z_0 = z_1\xi_1 + \cdots + z_\rho\xi_\rho$ where the z_i are in $\mathfrak{A}$ and ξ_i in $\mathfrak{K}$ are such that a sum $a_1\xi_1 + \cdots + a_\rho\xi_\rho = 0$ for the a_i in $\mathfrak{A}$ only when the a_i are all zero. Then if a is in $\mathfrak{A}$ we have $a \cdot z_0 - z_0 \cdot a = (a \cdot z_1 - z_1 \cdot a)\xi_1 + \cdots + (a \cdot z_\rho - z_\rho \cdot a)\xi_\rho = 0$, and $a \cdot z_i - z_i \cdot a = 0$. If also x is in $\mathfrak{A}$ we compute $a \cdot (x \cdot z_0) - (a \cdot x) \cdot z_0$ and other similar products, and see that the z_i are in $\mathfrak{Z}$, z_0 is in $\mathfrak{Z}_\mathfrak{K}$, $\mathfrak{Z}_0 = \mathfrak{Z}_\mathfrak{K}$.

An algebra $\mathfrak{A}$ over $\mathfrak{F}$ may be indecomposable but there may exist a scalar extension $\mathfrak{K}$ of $\mathfrak{F}$ such that $\mathfrak{A}_\mathfrak{K}$ is decomposable. Thus we call $\mathfrak{A}$ *absolutely*

indecomposable[7] if $\mathfrak{A}_{\mathfrak{K}}$ is indecomposable for every $\mathfrak{K}$. Moreover a decomposition (1) of a decomposable algebra will be called an *absolute decomposition* if the components $\mathfrak{A}_i$ are all absolutely indecomposable. Lemma 2 and the result above then imply

LEMMA 6. *An algebra $\mathfrak{A}$ is absolutely indecomposable if and only if its center $\mathfrak{Z}$ is absolutely indecomposable.*

LEMMA 7. *A decomposition* (1) *is an absolute decomposition if and only if the center of each component $\mathfrak{A}_i$ is absolutely indecomposable.*

We may also use Lemma 4 to obtain

LEMMA 8. *Let $\mathfrak{A}$ be an algebra with a unity quantity, $\mathfrak{K}$ and $\mathfrak{K}_0$ be scalar extensions of $\mathfrak{F}$ such that*

$$\mathfrak{A}_{\mathfrak{K}} = \mathfrak{A}_1 + \cdots + \mathfrak{A}_r, \qquad \mathfrak{A}_{\mathfrak{K}_0} = \mathfrak{B}_1 + \cdots + \mathfrak{B}_s \tag{5}$$

for absolutely indecomposable $\mathfrak{A}_i$ and $\mathfrak{B}_j$. Then $r = s$ and, if we imbed $\mathfrak{K}$ and $\mathfrak{K}_0$ in a scalar extension $\mathfrak{K}_1$ of $\mathfrak{F}$, there is a permutation $j_1, \cdots, j_r$ of $1, 2, \cdots, r$ such that $(\mathfrak{A}_i)_{\mathfrak{K}_1} = (\mathfrak{B}_{ji})_{\mathfrak{K}_1}$.

For $\mathfrak{A}_{\mathfrak{K}_1} = (\mathfrak{A}_{\mathfrak{K}})_{\mathfrak{K}_1} = (\mathfrak{A}_1)_{\mathfrak{K}_1} \oplus \cdots \oplus (\mathfrak{A}_r)_{\mathfrak{K}_1} = (\mathfrak{A}_{\mathfrak{K}_0})_{\mathfrak{K}_1} = (\mathfrak{B}_1)_{\mathfrak{K}_1} \oplus \cdots \oplus (\mathfrak{B}_r)_{\mathfrak{K}_1}$. Our result then follows from Lemma 4.

The center of a simple algebra $\mathfrak{A}$ with a unity quantity is a field and if separable is indecomposable only if its degree is one, $\mathfrak{A}$ is central. Thus we have

LEMMA 9. *A simple algebra with a unity quantity over $\mathfrak{F}$ and separable center is absolutely indecomposable if and only if it is central simple over $\mathfrak{F}$.*

4. Semi-simple algebras

We shall call an algebra $\mathfrak{A}$ over $\mathfrak{F}$ a *semi-simple algebra* if it has a unity quantity e and is the direct sum (1) of simple components $\mathfrak{A}_i$. Then we have seen in Lemmas 1, 2 that $\mathfrak{A}_i$ has a unity quantity e_i such that $e = e_1 + \cdots + e_r$, the center $\mathfrak{Z}_i$ of $\mathfrak{A}_i$ is a field, the center $\mathfrak{Z}$ of $\mathfrak{A}$ is the direct sum of the $\mathfrak{Z}_i$. We shall call $\mathfrak{A}$ *separable* if $\mathfrak{A}_{\mathfrak{K}}$ is semi-simple for every scalar extension $\mathfrak{K}$.

Let the center $\mathfrak{Z}$ of a simple algebra $\mathfrak{A}$ over $\mathfrak{F}$ and with a unity quantity e be a separable field. Then if $\mathfrak{K}$ is any scalar extension of $\mathfrak{F}$ the algebra $\mathfrak{Z}_{\mathfrak{K}} = \mathfrak{Z}_1 \oplus \cdots \oplus \mathfrak{Z}_r$, where $\mathfrak{Z}_i$ is a separable field over $\mathfrak{K}$ equivalent over $\mathfrak{F}$ to a composite of $\mathfrak{Z}$ and $\mathfrak{K}$. If e_i is its unity quantity the algebra $\mathfrak{Z}_i$ contains $\mathfrak{Z}e_i$ which is a field over $\mathfrak{F}$ equivalent over $\mathfrak{F}$ to $\mathfrak{Z}$ under the mapping $z \cdot e_i \to z$. We let $u_1, \cdots, u_s$ be a basis of $\mathfrak{A}$ over $\mathfrak{Z}$ and $u_g \cdot u_j = \sum_{k=1}^{s} u_k z_{gjk}$ for the z_{gjk} in $\mathfrak{Z}$ and see that $u_1 \cdot e_i, \cdots, u_s \cdot e_i$ are a basis of $\mathfrak{A}e_i$ over $\mathfrak{Z}e_i$, $(u_g \cdot e_i) \cdot (u_j \cdot e_i) = \sum_{k=1}^{s} (u_k \cdot e_i)(z_{gjk} \cdot e_i)$. Then we have a corresponding decomposition $\mathfrak{A}_{\mathfrak{K}} = \mathfrak{A}_1 + \cdots + \mathfrak{A}_r$ where $\mathfrak{A}_i = (\mathfrak{A}_{\mathfrak{K}})e_i = (\mathfrak{A}e_i)_{\mathfrak{K}}$. But the linear mapping $a \to a \cdot e_i$ is clearly an equivalence over $\mathfrak{F}$ of $\mathfrak{A}$ and $\mathfrak{A}e_i$, $\mathfrak{Z}e_i$ is the center of $\mathfrak{A}e_i$, $\mathfrak{A}_i$ is a simple algebra with center $\mathfrak{Z}_i$ over $\mathfrak{K}$.

Conversely let $\mathfrak{A}$ be simple and separable. If $\mathfrak{Z}$ is not separable it is known that there exists a scalar extension $\mathfrak{K}$ of $\mathfrak{F}$ such that $\mathfrak{Z}_{\mathfrak{K}}$ contains a quantity

[7] This too seems a desirable terminology.

$y \neq 0$, $y^h = 0$ for some positive integer h. But by hypothesis $\mathfrak{A}_\mathfrak{K} = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_r$ where the center of $\mathfrak{A}_\mathfrak{K}$ is a direct sum of fields and is a separable associative algebra $\mathfrak{Z}_\mathfrak{K}$. However y is properly nilpotent in $\mathfrak{Z}$, a contradicton. We have proved

LEMMA 10. *A simple algebra with a unity quantity is separable if and only if its center is a separable field.*

We also clearly have

LEMMA 11. *An algebra $\mathfrak{A}$ with a unity quantity is separable if and only if it is a direct sum of separable simple algebras.*

By a well known property of separable fields we have

LEMMA 12. *Let $\mathfrak{A}$ be separable. Then there exists a scalar extension $\mathfrak{K}$ such that $\mathfrak{A}_\mathfrak{K}$ is a direct sum $\mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_r$ of central simple algebras $\mathfrak{A}_i$ over $\mathfrak{K}$. This is an absolute decomposition of $\mathfrak{A}_\mathfrak{K}$ and r is the order above $\mathfrak{F}$ of the center of $\mathfrak{A}$.*

5. Extending groups of linear transformations

If $u_1, \cdots, u_n$ is a basis of $\mathfrak{A}$ over $\mathfrak{F}$, any linear transformation G over $\mathfrak{F}$ on $\mathfrak{A}$ has a matrix Γ such that G is given by $(\alpha_1 u_1 + \cdots + \alpha_n u_n)G = \beta_1 u_1 + \cdots + \beta_n u_n$ where

$$(\beta_1, \cdots, \beta_n) = (\alpha_1, \cdots, \alpha_n)\Gamma. \tag{6}$$

Then if $\mathfrak{K}$ is any scalar extension of $\mathfrak{F}$ we may indicate by $G_\mathfrak{K}$ the linear transformation on $\mathfrak{A}_\mathfrak{K}$ with the same matrix Γ. It is given by the equations above for the α_i and β_i now in $\mathfrak{K}$. Conversely if the matrix Γ of a linear transformation G_0 on $\mathfrak{A}_\mathfrak{K}$ with respect to a basis $u_1, \cdots, u_n$ of the original algebra $\mathfrak{A}$ has elements in $\mathfrak{F}$ then $G_0 = G_\mathfrak{K}$ where the matrix of G in $(\mathfrak{F})_n$ is also Γ.

As in the theory of *groups with operators* we shall consider algebras $\mathfrak{A}$ over $\mathfrak{F}$ with operator sets $\mathfrak{G}$ of linear transformations G over $\mathfrak{F}$. If $\mathfrak{K}$ is any scalar extension of $\mathfrak{F}$ we shall designate by $\mathfrak{G}_\mathfrak{K}$ the set of all G_K on $\mathfrak{A}_\mathfrak{K}$ for G in $\mathfrak{G}$.

A linear subspace $\mathfrak{B}$ over $\mathfrak{F}$ of $\mathfrak{A}$ will be called $\mathfrak{G}$-*allowable* if bG is in $\mathfrak{B}$ for every b of $\mathfrak{B}$ and G of $\mathfrak{G}$. Then a $\mathfrak{G}$-allowable ideal of $\mathfrak{A}$ will be called a $\mathfrak{G}$-ideal and we shall say that $\mathfrak{A}$ is $\mathfrak{G}$-*simple* if it has no $\mathfrak{G}$-ideals other than itself and the zero ideal. Finally we shall say that $\mathfrak{A}$ is $\mathfrak{G}$-*central* if every scalar extension $\mathfrak{A}_\mathfrak{K}$ of $\mathfrak{A}$ is $\mathfrak{G}_\mathfrak{K}$-simple.

We shall restrict all further attention to subgroups $\mathfrak{G}$ of the multiplicative group of all non-singular linear transformations on $\mathfrak{A}$ and shall call such a group $\mathfrak{G}$ an *extending*[8] *group* for $\mathfrak{A}$ if the unity quantity e of A has the property that $eG = e$ for every G of $\mathfrak{G}$. Then every subgroup of an extending group for $\mathfrak{A}$ is an extending group for $\mathfrak{A}$. Moreover $\mathfrak{G}$ is an extending group for $\mathfrak{A}$ if and only if $\mathfrak{G}_\mathfrak{K}$ is an extending group for $\mathfrak{A}_\mathfrak{K}$ where $\mathfrak{K}$ ranges over all scalar extensions of $\mathfrak{F}$. We now prove

THEOREM 1. *Let $\mathfrak{G}$ be an extending group for a semi-simple algebra $\mathfrak{A} = \mathfrak{A}_1 \oplus$*

---

[8] We shall use this terminology in our theorems so as to diminish the size of the statement of the hypotheses we shall find it necessary to make.

$\cdots \oplus \mathfrak{A}_r$ *with simple components* $\mathfrak{A}_i$, *and let there exist an* $a_i \neq 0$ *in* $\mathfrak{A}_1$ *and a transformation* G_i *in* $\mathfrak{G}$ *for each* $i = 1, \cdots, r$ *such that* a_iG_i *is in* $\mathfrak{A}_i$. *Then* $\mathfrak{A}$ *is* $\mathfrak{G}$*-simple, and is* $\mathfrak{G}$*-central if the* $\mathfrak{A}_i$ *are all central simple over* $\mathfrak{F}$.

For Lemma 5 states that every $\mathfrak{G}$-ideal $\mathfrak{B} = \mathfrak{B}_1 \oplus \cdots \oplus \mathfrak{B}_r$ where the intersection of $\mathfrak{B}$ and $\mathfrak{A}_i$ is the ideal $\mathfrak{B}_i$ of $\mathfrak{A}_i$. If $\mathfrak{B} \neq 0$ some $\mathfrak{B}_j \neq 0$, $\mathfrak{B}_j = \mathfrak{A}_j$ contains a_jG_j. But $\mathfrak{G}$ is a group and $\mathfrak{B}$ is a $\mathfrak{G}$-ideal only if $(a_jG_j)G_j^{-1} = a_j$ is in $\mathfrak{B}$. Hence a_j in $\mathfrak{A}_1$ is in $\mathfrak{B}_1$, $\mathfrak{B}_1 = \mathfrak{A}_1$. Then $\mathfrak{B}$ contains every a_i of our hypothesis and also every a_iG_i. These are non-zero quantities since $a_iG_i = 0$ implies that $a_iG_iG_i^{-1} = a_i = 0$. They are in $\mathfrak{B}$ and in $\mathfrak{A}_i$ and hence in $\mathfrak{B}_i$, $\mathfrak{B}_i \neq 0$, $\mathfrak{B}_i = \mathfrak{A}_i$, $\mathfrak{B} = \mathfrak{A}$ is $\mathfrak{G}$-simple. If every $\mathfrak{A}_i$ is central every $(\mathfrak{A}_i)_{\mathfrak{K}}$ is simple and our proof implies that $\mathfrak{A}_{\mathfrak{K}}$ is $\mathfrak{G}_{\mathfrak{K}}$-simple, $\mathfrak{A}$ is $\mathfrak{G}$-central.

THEOREM 2. *Let* $\mathfrak{H}$ *be a subgroup of an extending group* $\mathfrak{G}$ *for* $\mathfrak{A}$. *Then if* $\mathfrak{A}$ *is* $\mathfrak{H}$*-simple it is* $\mathfrak{G}$*-simple, and if* $\mathfrak{A}$ *is* $\mathfrak{H}$*-central it is* $\mathfrak{G}$*-central.*

The next result may be regarded as the trivial case $\mathfrak{H} = [I]$ of Theorem 2.

THEOREM 3. *A simple algebra* $\mathfrak{A}$ *is* $\mathfrak{G}$*-simple for every* $\mathfrak{G}$. *If* $\mathfrak{A}$ *is central simple it is* $\mathfrak{G}$*-central for every* $\mathfrak{G}$.

Every G of an extending group $\mathfrak{G}$ for an algebra $\mathfrak{A}$ induces a linear mapping $b \to bG$ of a linear subspace $\mathfrak{B}$ of $\mathfrak{A}$ on $\mathfrak{B}G$. If $\mathfrak{H}$ is any subgroup of $\mathfrak{G}$ such that $\mathfrak{B}H$ is a subset of $\mathfrak{B}$ for every H of $\mathfrak{H}$ then the mappings above are a group of non-singular linear transformations on $\mathfrak{B}$ induced by $\mathfrak{H}$. We shall use this terminology in the formulation of

THEOREM 4. *Let the center of a simple algebra* $\mathfrak{A}$ *over* $\mathfrak{F}$ *be a (separable) normal field* $\mathfrak{Z}$ *and let an extending group* $\mathfrak{G}$ *for* $\mathfrak{A}$ *have a subgroup inducing in* $\mathfrak{Z}$ *its automorphism group* $\mathfrak{H}$. *Then* $\mathfrak{A}$ *is* $\mathfrak{G}$*-central.*

For if $\mathfrak{B}$ is any $\mathfrak{G}_{\mathfrak{K}}$-ideal of $(\mathfrak{A})_{\mathfrak{K}}$ the set $\mathfrak{B}_{\mathfrak{N}}$ is a $\mathfrak{G}_{\mathfrak{N}}$-ideal of $\mathfrak{A}_{\mathfrak{N}}$, where $\mathfrak{N}$ is any scalar extension of $\mathfrak{F}$ containing $\mathfrak{K}$. But it is well known that $\mathfrak{N}$ may be so chosen that $\mathfrak{Z}_{\mathfrak{N}} = e_1\mathfrak{N} + \cdots + e_t\mathfrak{N}$ for pairwise orthogonal idempotents e_i, $\mathfrak{A}_{\mathfrak{N}} = \mathfrak{A}_i \oplus \cdots \oplus \mathfrak{A}_t$ as in Lemma 12, $\mathfrak{A}_i = (\mathfrak{A}_{\mathfrak{N}})e_i$ central simple. Moreover $e_i = e_1H_i$ for H_i in $\mathfrak{H}$ and hence $e_i = e_1G_i$ for G_i in $\mathfrak{G}$. By Theorem 1 $\mathfrak{A}_{\mathfrak{N}}$ is $\mathfrak{G}$-simple, $\mathfrak{B}_{\mathfrak{N}} = 0$ or $\mathfrak{A}_{\mathfrak{N}}$, $\mathfrak{B} = 0$ or $\mathfrak{A}_{\mathfrak{K}}$, $\mathfrak{A}$ is $\mathfrak{G}$-central.

COROLLARY I. *A normal field* $\mathfrak{A}$ *over* $\mathfrak{F}$ *is* $\mathfrak{G}$*-central with respect to its (extending) automorphism group* $\mathfrak{G}$.

Theorem 4 may be extended to direct sums and the result stated as

THEOREM 5. *Let* $\mathfrak{A}$ *be a* $\mathfrak{G}$*-simple algebra of Theorem* 1 *such that the center* $\mathfrak{Z}_i$ *of* $\mathfrak{A}_i$ *is a normal field over* $\mathfrak{F}$. *Then if* $\mathfrak{G}$ *has subgroups* $\mathfrak{G}_i$ *for* $i = 1, \cdots, r$ *such that* $\mathfrak{G}_i$ *induces in* $\mathfrak{Z}_i$ *its automorphism group, the algebra* $\mathfrak{A}$ *is* $\mathfrak{G}$*-central.*

This generalization is a corollary of Theorem 4 and the following

THEOREM 6. *Let* $\mathfrak{A}$ *be a* $\mathfrak{G}$*-simple algebra of Theorem* 1 *and* $\mathfrak{G}$ *have subgroups* $\mathfrak{G}_i$ *inducing in* $\mathfrak{A}_i$ *an extending group* $\mathfrak{H}$ *such that* $\mathfrak{A}_i$ *is* $\mathfrak{G}_i$*-central for every* $i = 1, \cdots, r$. *Then* $\mathfrak{A}$ *is* $\mathfrak{G}$*-central.*

To prove this result we see that every $\mathfrak{G}$-ideal of $\mathfrak{A}$ has the form $\mathfrak{B} = \mathfrak{B}_1 + \cdots \oplus \mathfrak{B}_r$ where $\mathfrak{B}_i$ is the ideal of $(\mathfrak{A}_i)_{\mathfrak{K}}$ which is the intersection of $\mathfrak{B}$ and $(\mathfrak{A}_i)_{\mathfrak{K}}$. Then $\mathfrak{B}_i(\mathfrak{H}_i)_{\mathfrak{K}}$ is in $\mathfrak{B}$ and in $(\mathfrak{A}_i)_{\mathfrak{K}}$ and is in $\mathfrak{B}_i$, $\mathfrak{B}_i$ is an $(\mathfrak{H}_i)_{\mathfrak{K}}$-ideal. Since $\mathfrak{A}_i$ is central $\mathfrak{B}_i = 0$ or $(\mathfrak{A}_i)_{\mathfrak{K}}$. The remainder of our proof is exactly as in the proof of Theorem 1.

6. Crossed extensions

A quantity g of an algebra $\mathfrak{A}$ has been called *non-singular* if it is neither a right nor a left-divisor of zero. Then the right multiplication R_g and the left multiplication L_g are non-singular and we have

LEMMA 13. *Let a and g be quantities of an algebra $\mathfrak{A}$ such that g is non-singular. Then if an ideal B of A contains either $a \cdot g$ or $g \cdot a$ it contains a.*

For $a \cdot g = aR_g$ is in $\mathfrak{B}$ and so is $(aR_g) \cdot g = a(R_g)^2$. If $a(R_g)^k$ is in $\mathfrak{B}$ so is $[a(R_g)^k] \cdot g = a(R_g)^{k+1}$. Hence $\mathfrak{B}$ contains $a(R_g)^k$ for every positive integer k. Since $\mathfrak{B}$ is a linear space it contains every $a\phi(R_g)$ for $\phi(R_g) = \alpha_1 R_g + \alpha_2(R_g)^2 + \cdots + \alpha_r(R_g)^r$ and the α_i in $\mathfrak{F}$. But the constant term β_n of the characteristic function $\psi(\lambda) = \lambda^n + \beta_1\lambda^{n-1} + \cdots + \beta_n$ of R_g is not zero, $\psi(R_g) = 0$, the identity transformation $I = -\beta_n^{-1}[\beta_{n-1}R_g + \cdots + (R_g)^n]$ is $a\phi(R_g)$, $\mathfrak{B}$ contains $aI = a$. Similarly if $\mathfrak{B}$ contains $g \cdot a$ it contains every $a\phi(L_g)$ and a.

We now make the

DEFINITION. *Let $\mathfrak{A}$ be an algebra with a unity quantity e and $\mathfrak{G}$ be an extending group for $\mathfrak{A}$. Then a set*

$$\mathfrak{g} = \{g_{S,T}\} \tag{7}$$

of quantities $g_{S,T}$ of $\mathfrak{A}$ will be called an extending set[9] *for $\mathfrak{A}$ by $\mathfrak{G}$ if g contains one and only one $g_{S,T}$ for every pair of transformations S and T of $\mathfrak{G}$, the $g_{S,T}$ are non-singular quantities of $\mathfrak{A}$, and*

$$g_{I,S} = g_{S,I} = e, \tag{8}$$

for every S of $\mathfrak{G}$.

We shall now proceed to our definition of new classes of algebras. We let $\mathfrak{A}$ be any algebra with a unity quantity e, n be the order over $\mathfrak{F}$ of $\mathfrak{A}$, $\mathfrak{G}$ be a finite[10] extending group of order m for $\mathfrak{A}$, $\mathfrak{g}$ be an extending set for $\mathfrak{A}$ by $\mathfrak{G}$. We let $\mathfrak{H}$ be a subset of $\mathfrak{G}$ containing the identity transformation. Then we shall define an algebra

$$\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g}), \tag{9}$$

of order $\nu = nm$ over $\mathfrak{F}$, which we shall call the *crossed extension* of $\mathfrak{A}$ by $\mathfrak{H}$ and $\mathfrak{G}$ with extension set $\mathfrak{g}$. We shall also call the integer m the *extension index* of $\mathfrak{A}$ under $\mathfrak{C}$.

We let $\mathfrak{N}$ be a linear space of order ν over $\mathfrak{F}$ so that $\mathfrak{N}$ is the supplementary sum

$$\mathfrak{N}_1 + \cdots + \mathfrak{N}_m, \tag{10}$$

---

[9] The term extending set is preferable to that of factor set which we reserve for extending sets restricted so that the algebras we construct will be associative.

[10] It seems clear that if we take $\mathfrak{G}$ to be an infinite group our construction will be valid if we take the corresponding linear space $\mathfrak{N}$ to consist of vectors with finitely many non-zero coordinates. Moreover it seems that our hypotheses insuring that the result is a simple algebra will be also sufficient for the algebras of infinite order. It would be interesting to take the case where $\mathfrak{G}$ consists of the non-zero quantities of a division algebra, as well as that where $\mathfrak{N}$ is a Hilbert space.

of linear subspaces $\mathfrak{N}_i$ each of order n over $\mathfrak{F}$. Then there exist corresponding non-singular linear transformations $C_1, \cdots, C_m$ in $(\mathfrak{F})$, such that $C_1 = I$, is the identity transformation on $\mathfrak{N}$,

$$\mathfrak{N}_i = \mathfrak{N}_1 C_i \qquad (i = 1, \cdots, m). \tag{11}$$

Thus every quantity of $\mathfrak{N}$ is uniquely expressible in the form

$$a = a_1 C_1 + \cdots + a_m C_m \qquad (a_i \text{ in } \mathfrak{N}_1). \tag{12}$$

Observe next that the linear spaces $\mathfrak{A}$ and $\mathfrak{N}$ have the same order and thus that it is possible to take $\mathfrak{A} = \mathfrak{N}_1$. We do this and have thus imbedded the algebra $\mathfrak{A}$ in $\mathfrak{N}$ as a linear subspace of $\mathfrak{N}$. We shall actually define $\mathfrak{C}$ to be an algebra whose quantities are the vectors of $\mathfrak{N}$ and we shall formulate our definition so that $\mathfrak{A}$ will be a subalgebra of $\mathfrak{C}$.

Let us order the transformations of $\mathfrak{G}$ in any order such that the first transformation is I and thus have the notations

$$S = I, S_2, \cdots, S_m \tag{13}$$

for these transformations. We have then defined a one-to-one mapping

$$S_i \rightarrow C_i = C_{S_i} \tag{14}$$

of $\mathfrak{G}$ on the set of C_i such that $C_1 = C_I = I_\nu$, the identity transformation on the space of order ν. If $S = S_i$ we designate by a_S the coefficient a_i of $C_i = C_S$ and may thus write every α of $\mathfrak{N}$ uniquely in the form

$$\alpha = \sum_S a_S C_S, \tag{15}$$

for the a_S in $\mathfrak{A}$ where the sum is taken over all S of $\mathfrak{G}$. Write

$$\kappa = \sum_T x_T C_T, \tag{16}$$

and define

$$\alpha \cdot \kappa = \sum_U y_U C_U,$$

for the x_T and y_U in $\mathfrak{A}$ where the y_U are to to be determined. Then the distributive law holds only if

$$\alpha \cdot \kappa = \sum_{S,T} (a_S C_S) \cdot (x_T C_T). \tag{17}$$

We now let

$$a_S C_S \cdot x_T C_T = y_{S,T} C_{ST}, \tag{18}$$

so that if we write $U = ST$, $T = S^{-1}U$, then we have

$$y_U = \sum_S y_{S,S^{-1}U}. \tag{19}$$

We shall then complete our definition when we express the $y_{S,T}$ in terms of terms of a_S, x_T and $\mathfrak{g}$, and we do this by defining the function

$$w(S, T, a, x) = aT\cdot x \qquad (S \text{ in } \mathfrak{H}), \tag{20}$$

$$w(S, T, a, x) = x\cdot aT \qquad (S \text{ not in } \mathfrak{H}), \tag{21}$$

for every ordered pair a, x of quantities of $\mathfrak{A}$ where the products indicated are *products in* $\mathfrak{A}$ of its quantities. We are then able to write the desired formulas

$$y_{S,T} = w(ST, I, g_{S,T}, w_{S,T}), \qquad w_{S,T} = w(S, T, a_S, x_T). \tag{22}$$

In particular $e = g_{I,T} = g_{S,I}$ for every S and hence we have

$$y_{I,T} = w_{I,T} = a_I T\cdot x_T, \qquad y_{S,I} = w_{S,I} = w(S, I, a_S, x_I). \tag{23}$$

Conversely let $y_{S,T}$ be defined by (20), (21), (22), so that the y_U are defined uniquely by (19). Since $\mathfrak{A}$ is a linear algebra it is clear that the $y_{S,T}$ are linear in the x_T and thus also in the coordinates of x, the y_U are linear in x, $a\cdot x$ is linear in x. Also every T is a linear transformation, the $y_{S,T}$ are linear in $a_S T$ and hence in a_S, $a\cdot x$ is linear in a. It follows that $\mathfrak{E}$ is a linear algebra.

We note that $y_{S,T} = 0$ if either a_S or x_T is zero. Then if α and κ are in $\mathfrak{A}$ we have $\alpha = a_I$, $\kappa = x_I$, all the $y_{S,T}$ are zero except $y_{I,I} = a_I\cdot x_I$ by (23), $\mathfrak{A}$ is a subalgebra of $\mathfrak{E}$. In fact we may prove

THEOREM 7. *The algebra* $\mathfrak{E}$ *of* (7)–(21) *contains* $\mathfrak{A}$ *as a subalgebra and the unity quantity* e *of* $\mathfrak{A}$ *is the unity quantity of* $\mathfrak{E}$. *Every quantity of* $\mathfrak{E}$ *is uniquely expressible in the form*

$$\alpha = \sum_S u_S\cdot a_S \qquad (S \textit{ in } G,\ a_S \textit{ in } \mathfrak{A}), \tag{24}$$

where $u_S = eC_S$, $u_I = e$. *The quantities* u_S *are non-singular quantities of* E *such that* $u_S\cdot a_S = a_S C_S$ *and*

$$u_S\cdot u_T = u_{ST} g_{S,T}, \tag{25}$$

for every S *and* T *of* $\mathfrak{G}$. *Then the definitive properties of* E *are completely given by* (20), (21), (25) *and*

$$(u_S\cdot a_S)\cdot(u_T\cdot x_T) = (u_S\cdot u_T)\cdot[w(S, T, a_S, x_T)]. \tag{26}$$

For by (23) we have $y_{I,T} = x_T = y_T$ if $\alpha = e$, $e\cdot\kappa = \kappa$. Similarly $y_{S,I} = w(S, I, a_S, e) = a_S$ if $\kappa = e$, $y_{S,T} = 0$ unless $S^{-1}U = T = I$, $S = U$. Then $y_U = y_{U,I} = a_U$, $\eta = \alpha = \alpha\cdot e$, e is the unity quantity of $\mathfrak{E}$. Now $u_S\cdot a_S = eC_S\cdot a_S C_I = y_{S,I}C_S = [w(S, I, e, a_S)]C_S = a_S C_S$ so that (15) is equivalent to (24). Also (18) becomes

$$(u_S\cdot a_S)\cdot(u_T\cdot x_T) = u_{ST}\cdot y_{S,T}. \tag{27}$$

But (25) follows from (27) if we put $a_S = x_T = e$ and use the property that $eT = e$. The definition (22) then states that (27) is equivalent to

$$(u_S\cdot a_S)\cdot(u_T\cdot x_T) = u_{ST}\cdot[w(ST, I, g_{S,T}, w_{S,T})]. \tag{28}$$

But $(u_S \cdot u_T) \cdot w_{S,T} = (u_{ST} \cdot g_{ST}) \cdot (u_I \cdot w_{S,T}) = u_{ST} \cdot [w(ST, I, g_{S,T}, w_{S,T})]$ and (28) implies (26). Conversely (26) and (25) imply that $(u_S \cdot a_S) \cdot (u_T \cdot x_T) = (u_{ST} \cdot g_{S,T}) \cdot w_{S,T} = (u_{ST} \cdot g_{S,T}) \cdot (u_I w_{S,T})$ and the fact that $u_{ST} \cdot u_I = u_{ST}$ used with (26) implies (27).

Since $g_{S,T}$ is non-singular we have $u_S \cdot (u_T \cdot x_T) = u_{ST} \cdot y_{S,T}$ where $y_{S,T} = g_{S,T} \cdot x_T$ or $x_T \cdot g_{S,T}$ is not zero unless $x_T = 0$. Then $u_S \cdot \kappa = \sum_T u_{ST} \cdot y_{ST} \neq 0$ unless $\kappa = 0$, u_S is not a left divisor of zero. Similarly u_S is not a right divisor of zero and is a non-singular quantity of $\mathfrak{A}$. This proves our theorem.

7. A non-simple crossed extension

The crossed extension $\mathfrak{C}$ need not be simple even when $\mathfrak{A}$ is $\mathfrak{G}$-simple, nor need $\mathfrak{C}$ be central when $\mathfrak{A}$ is $\mathfrak{G}$-central. Let us give an example here of such an algebra. We take $\mathfrak{F}$ to be a field of real numbers, $\mathfrak{A} = \mathfrak{A}_1 \oplus \mathfrak{A}_2$ where e_1 and e_2 are the respective unity quantities of the quadratic fields $\mathfrak{A}_1$ and $\mathfrak{A}_2$ defined by

$$u_1^2 = -e_1, \qquad u_2^2 = -e_2, \qquad \mathfrak{A}_1 = e_1\mathfrak{F} + u_1\mathfrak{F}, \qquad \mathfrak{A}_2 = e_2\mathfrak{F} + u_2\mathfrak{F}.$$

Then every quantity of $\mathfrak{A}$ is uniquely expressible in the form

$$a = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 \qquad (\alpha_i \text{ in } \mathfrak{F}), \tag{29}$$

where $u_3 = e_1 - u_1$, $u_4 = e_2 - u_2$. We let $\mathfrak{G}$ be the group of linear transformations on $\mathfrak{A}$ obtained by applying the permutations (13), (24), (13) (24), (12) (34), (14) (23), (1432), (1234) and the identity to the subscripts i on the u_i in (29). Then $\mathfrak{G}$ is an extending group of order eight for $\mathfrak{A}$ and is known[11] to be generated by the transformation S obtained by applying (13) and the transformation P obtained by applying (12) (34). We let T be the transformation obtained by applying (24) and have

$$ST = TS, \qquad S^2 = T^2 = I, \qquad SP = PT, \qquad PS = TP. \tag{30}$$

Here TP is the transformation obtained by applying the cycle (1234) to the subscripts of the basal quantities in (29), and PT is its inverse obtained by applying (4321).

The algebra $\mathfrak{A}$ has the property that $\mathfrak{A}_{\mathfrak{K}} = \mathfrak{B}_1 \oplus \mathfrak{B}_2 \oplus \mathfrak{B}_3 \oplus \mathfrak{B}_4$ where $\mathfrak{B}_i$ has order one over the field $\mathfrak{K}$ of all complex numbers, $\mathfrak{B}_i = v_i\mathfrak{F}$ such that $2v_1 = u_3 + (1 + i)\, u_1$, $2v_3 = u_3 + (1 - i)\, u_1$, $2v_2 = u_4 + (1 + i)\, u_2$, $2v_4 = u_4 + (1 - i)\, u_2$ are pairwise orthogonal idempotents. Any non-zero ideal $\mathfrak{B}$ of $\mathfrak{A}_{\mathfrak{K}}$ contains one of the v_i. If $\mathfrak{B}$ is a $\mathfrak{G}_{\mathfrak{K}}$-ideal it contains with v_1 the quantity u_3 and then we apply P and its powers to get all the u_i, $\mathfrak{B} = \mathfrak{A}_{\mathfrak{K}}$. If $\mathfrak{B}$ contains v_3 and hence $2u_1 + u_3$ we apply S to get $2u_3 + u_1$ and hence the quantity $2(2u_1 + u_3) - (2u_3 + u_1) = 3u_1$. Again $\mathfrak{B} = \mathfrak{A}_{\mathfrak{K}}$. Similarly if $\mathfrak{B}$ contains either v_2 or v_4 it is equal to $\mathfrak{A}_{\mathfrak{K}}$, $\mathfrak{A}_{\mathfrak{K}}$ is $\mathfrak{G}$-simple, $\mathfrak{A}$ is $\mathfrak{G}$-central.

Observe that S carries quantities of $\mathfrak{A}_1$ into other quantities and leaves e_1 as well as every quantity of $\mathfrak{A}_2$ unaltered, T carries quantities of $\mathfrak{A}_2$ into other

[11] cf. L. E. Dickson, *Modern Algebraic Theories*, pp. 145–6.

quantities such that e_2 is unaltered and T leaves all quantities of $\mathfrak{A}_1$ unaltered. We form the crossed extension $\mathfrak{C}$ defined above for all quantities in the extension set equal to the unity quantity of $\mathfrak{A}$ and for $\mathfrak{H}$ the identity group. Let $\mathfrak{z} = u_S \cdot e_2 + u_T \cdot e_1$. Now $\mathfrak{z} \cdot a = u_S \cdot (a_2 \cdot e_2) + u_T \cdot (a_1 \cdot e_1)$ for every $a = a_1 + a_2$ such that a_1 is in $\mathfrak{A}_1$ and a_2 in $\mathfrak{A}_2$. Also $a \cdot \mathfrak{z} = u_S \cdot (a_S \cdot e_2) + u_T \cdot (aT \cdot e_1) = \mathfrak{z} \cdot a$ since $(aS) \cdot e_2 = (a_1 S + a_2) \cdot e_2 = a_2 \cdot e_2$, $aT \cdot e_1 = (a_2 S + a_1) \cdot e_1 = a_1 \cdot e_1$. Moreover $u_S \cdot \mathfrak{z} = u_{ST} \cdot e_1 + e_2 = \mathfrak{z} \cdot u_S = u_{TS} \cdot e_1 + e_2$, $u_T \cdot \mathfrak{z} = u_{TS} \cdot e_2 + e_1 = \mathfrak{z} \cdot u_T$. Finally $u_P \cdot \mathfrak{z} = u_{PS} \cdot e_2 + u_{PT} \cdot e_1$, $\mathfrak{z} \cdot u_P = u_{SP} \cdot e_1 + u_{TP} \cdot e_2 = u_P \cdot \mathfrak{z}$ by (30). That $(\mathfrak{a} \cdot \mathfrak{x}) \cdot \mathfrak{y} = \mathfrak{a} \cdot (\mathfrak{x} \cdot \mathfrak{y})$ when one of the factors is $\mathfrak{z}$ follows from the fact that $\mathfrak{A}$ is a commutative associative algebra and that $u_Q \cdot u_R = u_{QR}$. Then $\mathfrak{z}$ is in the center of our crossed extension $\mathfrak{C}$. But $\mathfrak{z}^2 = (u_S \cdot e_2)^2 + (u_T \cdot e_1)^2 + (u_S \cdot e_2)(u_T \cdot e_1) + (u_T \cdot e_1)(u_S \cdot e_2) = e$ since $(u_S \cdot e_2)^2 = e_2$, $(u_T \cdot e_1)^2 = e_1$, and the other two terms are equal to $u_{ST} \cdot (e_1 \cdot e_2) = 0$. It follows that $\mathfrak{C}$ is a direct sum $\mathfrak{C} = \mathfrak{C}\mathfrak{z}_1 \oplus \mathfrak{C}\mathfrak{z}_2$, $2\mathfrak{z}_1 = e - \mathfrak{z}$, $2\mathfrak{z}_2 = e + \mathfrak{z}$, $\mathfrak{C}$ is neither simple nor central.

8. Simple crossed extensions

We have just seen that $\mathfrak{A}$ may be $\mathfrak{G}$-simple but its extension $\mathfrak{C}$ not a simple algebra. Thus we shall have to make additional hypotheses if we wish every $\mathfrak{C}$ defined for the given $\mathfrak{A}$, $\mathfrak{G}$, $\mathfrak{H}$ to be simple. These conditions are really a part of the usual associative crossed product definition, but are hidden in the more explicit and special nature of those algebras.

Let us call a linear transformation S on the linear space $\mathfrak{A}$ an *inner*[12] or an *outer* transformation for this algebra according as there is or is not a quantity $b \neq 0$ in $\mathfrak{A}$ such that

$$b \cdot x = xS \cdot b. \tag{31}$$

The identity transformation I is one of a class of inner transformations on $\mathfrak{A}$ which have the property that (31) holds for b in the center of $\mathfrak{A}$. We shall call any such transformation a *semi-identity* transformation for $\mathfrak{A}$. If S is semi-identical and $\mathfrak{A}$ is semi-simple we may write $b = b_1 + \cdots + b_s$, $\mathfrak{A} = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_s \oplus C$ for simple algebras $\mathfrak{A}_i$ such that $b_i \neq 0$ is in the center of $\mathfrak{A}_i$ and (31) becomes $b \cdot (x - xS) = 0$. But there exist $d_1, \cdots, d_s$ in the center of $\mathfrak{A}_1, \cdots, \mathfrak{A}_s$ such that $d_i \cdot b_i = e_i$, $f = e_1 + \cdots + e_s$ is the unity quantity of the ideal $\mathfrak{B} = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_s$ of $\mathfrak{A}$, $f \cdot (x - xS) = 0$. Then $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{C}$ such that $y - yS$ is in $\mathfrak{C}$ for every y of $\mathfrak{B}$, $\mathfrak{C}S = \mathfrak{C}$, $b \cdot x = xS \cdot b$ for every b in the center of $\mathfrak{B}$.

We now have a terminology for the hypotheses we shall require, and we shall prove

THEOREM 8. *Let $\mathfrak{A}$ be a semi-simple algebra, $\mathfrak{G}$ be an extending group for $\mathfrak{A}$ such that $\mathfrak{A}$ is $\mathfrak{G}$-simple and I is the only semi-identity transformation for $\mathfrak{A}$ in $\mathfrak{G}$.*

---

[12] If $\mathfrak{S}$ is an automorphism it is inner in the ordinary sense only when the quantity b is non-singular. However we shall require the (only slightly) more general hypothesis we give here.

Then if there is any subset $\mathfrak{H}$ of $\mathfrak{G}$ such that $\mathfrak{H}$ consists of I and outer[13] *transformations for $\mathfrak{A}$ the crossed extensions $\mathfrak{E} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ are simple algebras.*

For let $\rho(\mathfrak{a})$ be the number of non-zero coefficients a_S in the unique expression (24) of any $\mathfrak{a}$ in $\mathfrak{E}$ for a_S in $\mathfrak{A}$. Then $\rho(0) = 0$, $\rho(\mathfrak{a})$ is a positive integer for every $\mathfrak{a} \neq 0$. Let $\mathfrak{B}$ be a non-zero ideal of $\mathfrak{E}$ and ρ be the least $\rho(\mathfrak{y})$ for any $\mathfrak{y} \neq 0$ in $\mathfrak{B}$. Then some $\mathfrak{b}$ in $\mathfrak{B}$ has the property $\rho(\mathfrak{b}) = \rho$ and if we write $\mathfrak{b} = \sum u_S \cdot b_S$ as in (24) there is an S_0 such that $b_{S_0} \neq 0$. Then $u_T \cdot \mathfrak{b} = \sum_S u_{TS} \cdot c_{TS}$ where c_{TS} is the product of b_S by a non-singular quantity $g_{T,S}$ of $\mathfrak{A}$ and is not zero if $b_S \neq 0$. Take $T = S_0^{-1}$ and have a quantity $\mathfrak{c}$ in $\mathfrak{B}$ such that $\rho(\mathfrak{c}) = \rho$, $\mathfrak{c} = \sum_S u_S \cdot c_S$, $c_I \neq 0$. We now let $\mathfrak{D}$ be the set of all finite sums of terms of the form $(x \cdot \mathfrak{c}) \cdot y$, $x \cdot (y \cdot \mathfrak{c})$, for x and y in $\mathfrak{A}$. It follows that every $\mathfrak{d}$ of $\mathfrak{D}$ has the form

$$\mathfrak{d} = d_I + \sum_{i=2}^{\rho} u_{S_i} d_{S_i},$$

for a fixed set $I, S_2, \cdots, S_\rho$ in $\mathfrak{G}$ and with the d_S in $\mathfrak{A}$. Then $\mathfrak{c}$ is in the set $\mathfrak{D}$ and $\mathfrak{D}$ is a non-zero linear subspace of the ideal $\mathfrak{B}$ such that $\rho(\mathfrak{d}) = \rho$ for every non-zero $\mathfrak{d}$ of $\mathfrak{D}$. Moreover the quantities d_I consist of all finite sums of the form $(x \cdot c_I) \cdot y$ or $x \cdot (c_I \cdot y)$ for x and y in $\mathfrak{A}$, the set $\mathfrak{N}$ of all the d_I is a non-zero ideal of $\mathfrak{A}$. Leti f_I be its unity quantity so that f_I is in the center of $\mathfrak{A}$ and there is a quantty $\mathfrak{f}$ in $\mathfrak{D}$ such that

$$\mathfrak{f} = f_I + \sum_{i=2}^{\rho} u_{S_i} f_{S_i} \qquad (f_{S_i} \text{ in } \mathfrak{A}).$$

If $\rho < 1$ and some $T = S_i$ is not in $\mathfrak{H}$ it is not a semi-identical transformation and there exists a quantity x in $\mathfrak{A}$ such that $h_I = x \cdot f_I - f_I \cdot xT = f_I \cdot (x - xT) \neq 0$. The corresponding $\mathfrak{h} = x \cdot \mathfrak{f} - \mathfrak{f} \cdot xT \neq 0$ is in $\mathfrak{D}$ so that $\rho(\mathfrak{h}) = \rho$. But the term of $\mathfrak{h}$ involving u_T is $x \cdot (u_T \cdot f_T) - (u_T \cdot f_T) \cdot xT = u_T \cdot (xT \cdot f_T - xT \cdot f_T) = 0$, a contradiction. Hence every S_i is in $\mathfrak{H}$ and if $\rho > 1$ there is an $S_i = T$ which is an outer transformation, there must exist a quantity x in $\mathfrak{A}$ such that $h_T = xT \cdot f_T - f_T \cdot x \neq 0$, $x \cdot (u_T \cdot f_T) - (u_T \cdot f_T) \cdot x = u_T \cdot h_T \neq 0$ is the term involving u_T in $\mathfrak{h} = x \cdot \mathfrak{f} - \mathfrak{f} \cdot x$. Then $\mathfrak{h} \neq 0$ is in $\mathfrak{D}$ and $\rho(\mathfrak{h}) = \rho$. However f_I is in the center of $\mathfrak{A}$ and $h_I = x \cdot f_I - f_I \cdot x = 0$, a contradiction.

This proves that $\rho = 1$ and that the intersection $\mathfrak{U}$ of $\mathfrak{B}$ and $\mathfrak{A}$ is a non-zero ideal of $\mathfrak{A}$. If $\mathfrak{U}$ contains $\mathfrak{d}$ and S is in $\mathfrak{G}$ we write $T = S^{-1}$ and have $u_T \cdot (\mathfrak{d} \cdot u_S) = \mathfrak{h}$ where $\mathfrak{h} = g_{T,S} \cdot \mathfrak{d}S$ or $\mathfrak{d}S \cdot g_{T,S}$ is in $\mathfrak{U}$. By Lemma 13 so is dS. Thus $\mathfrak{U}$ is a $\mathfrak{G}$-ideal of $\mathfrak{A}$ and $\mathfrak{U} = \mathfrak{A}$ since $\mathfrak{A}$ is $\mathfrak{G}$-simple. Then $\mathfrak{B}$ contains e and is the unit ideal $\mathfrak{E}$, $\mathfrak{E}$ is a simple algebra.

We note that the example in Section 6 of a crossed extension which is not a simple algebra failed to satisfy our hypotheses precisely in that $\mathfrak{G}$ contained semi-identical transformations $S \neq I$.

We shall not try to compute the center of a crossed extension $\mathfrak{E}$ and so to

[13] In the case of ordinary crossed products $\mathfrak{G} = \mathfrak{H}$ and $\mathfrak{A}$ is a field. Our hypotheses are then satisfied.

prove that a given $\mathfrak{C}$ is central simple but we shall rather try to see that our hypotheses are in a form such that they hold also when the field $\mathfrak{F}$ is extended.[14] Thus we prove

LEMMA 14. *A linear transformation S on a separable algebra $\mathfrak{A}$ is a semi-identical transformation or an inner transformation for $\mathfrak{A}$ if and only if $S_{\mathfrak{K}}$ has the corresponding property for $\mathfrak{A}_{\mathfrak{K}}$, where $\mathfrak{K}$ is any scalar extension of $\mathfrak{F}$.*

For if (31) holds for a quantity b in $\mathfrak{A}$ and every x of $\mathfrak{A}$ we will also have $b\cdot x = xS_{\mathfrak{K}}\cdot b$ for every x in $\mathfrak{A}_{\mathfrak{K}}$, $S_{\mathfrak{K}}$ is inner when S is. Also $S_{\mathfrak{K}}$ is semi-identical when S is since if $\mathfrak{Z}$ is the center of $\mathfrak{A}$ the center of $\mathfrak{A}_{\mathfrak{K}}$ is $\mathfrak{Z}_{\mathfrak{K}}$. Conversely let $y\cdot x = xS_{\mathfrak{K}}\cdot y$ for every x of $\mathfrak{A}_{\mathfrak{K}}$ and a fixed quantity y in $\mathfrak{A}_{\mathfrak{K}}$. Then we may write $y = y_1\xi_1 + \cdots + y_s\xi_s$ for y_j in $\mathfrak{A}$ where the ξ_j are in $\mathfrak{K}$ and are such that $a_1\xi_1 + \cdots + a_s\xi_s = 0$ for a_i in $\mathfrak{A}$ if and only if the a_i are all zero. Take x in $\mathfrak{A}$ and so obtain $xS_{\mathfrak{K}} = xS$ in $\mathfrak{A}$, $y\cdot x - xS_{\mathfrak{K}}\cdot y = \sum (y_j\cdot x - xS_{\mathfrak{K}}\cdot y_j)\xi_j = 0$, $y_j\cdot x = xS\cdot y$. Hence if $S_{\mathfrak{K}}$ is inner so is S. If $S_{\mathfrak{K}}$ is semi-identical the quantity y may be taken to be in $\mathfrak{Z}_{\mathfrak{K}}$, y_1 is in $\mathfrak{Z}$, S is semi-identical.

We may thus apply Lemma 14 to Theorem 8 and obtain

THEOREM 9. *Let $\mathfrak{A}$ be a separable algebra,[15] $\mathfrak{G}$ be an extending group for $\mathfrak{A}$ such that $\mathfrak{A}$ is $\mathfrak{G}$-central and I is the only semi-identity transformation for $\mathfrak{A}$ in $\mathfrak{G}$. Then if $\mathfrak{H}$ is any subset of $\mathfrak{G}$ consisting of I and outer transformations for $\mathfrak{A}$ the crossed extensions $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ are central simple algebras.*

If $\mathfrak{A}$ is a simple algebra the only semi-identity transformation for $\mathfrak{A}$ is I and we have

THEOREM 10. *Let $\mathfrak{G}$ be an extending group for a simple algebra $\mathfrak{A}$, $\mathfrak{H}$ consist of I and outer transformations for $\mathfrak{A}$ in $\mathfrak{G}$. Then every crossed extension of $\mathfrak{A}$ is a simple algebra.*

If $\mathfrak{H}$ consists of I alone we shall write

$$\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$$

for the corresponding crossed extensions. These are surely the most interesting of our new algebras and we shall state the results of Theorems 8 and 9 for such algebras as

THEOREM 11. *Let $\mathfrak{G}$ be an extending group for a simple algebra $\mathfrak{A}$. Then every crossed extension $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$ is a simple algebra. Moreover if $\mathfrak{A}$ is central simple so is $\mathfrak{C}$.*

9. Associativity

A crossed extension $\mathfrak{C}$ is associative if and only if $\mathfrak{A}$ is associative and $[(u_S\cdot a_S)(u_T\cdot x_T)]\cdot u_P\cdot w_P = (u_S\cdot a_S)\cdot[(u_T\cdot x_T)\cdot(u_P\cdot w_P)]$ for every a_S, x_T, w_P of $\mathfrak{A}$ and S, T, P of $\mathfrak{G}$. If $\mathfrak{G} \neq \mathfrak{H}$ we take S not in $\mathfrak{H}$, $a_S = e$, $T = P = I$ and have $u_S\cdot(x\cdot w) = (u_S\cdot x)\cdot w = u_S(w\cdot x)$ which is possible in an associative

[14] This seems to be the best possible method of proof even for ordinary crossed products.

[15] In the associative case the simple components of $\mathfrak{A}$ are necessarily equivalent, since $\mathfrak{A}$ is $\mathfrak{G}$-simple and $\mathfrak{G} = \mathfrak{H}$ is composed of automorphisms. However this does not appear to be necessary here and this question should be studied.

algebra $\mathfrak{C}$ if and only if $\mathfrak{A}$ is commutative. But when $\mathfrak{A}$ is commutative the algebras $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ are the same for every $\mathfrak{H}$ and we may take $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$. Hence we have $\mathfrak{H} = [I]$ in every case. We now compute $(u_S \cdot a_S) \cdot (u_T \cdot x_T) = u_{ST} \cdot (g_{S,T} \cdot a_S T \cdot x_T)$. Similarly $(u_T \cdot x_T) \cdot (u_P \cdot w_P) = u_{TP} \cdot (g_{T,P} \cdot x_T P \cdot w_P)$. Multiply the first of these products on the right by $u_P \cdot w_P$ and the second on the left by $u_S \cdot a_S$. The resulting products each have the non-singular left factor u_{STP}, and if $\mathfrak{C}$ is associative we have

$$g_{ST,P}[g_{S,T} \cdot (a_S T \cdot x_T)]P \cdot w_P = g_{S,TP} \cdot (a_S TP) \cdot [g_{T,P} \cdot (x_T P \cdot w_P)]. \tag{32}$$

Put $a_S = x_T = w_P$ equal to the unity quantity e of $\mathfrak{A}$ and obtain

$$g_{ST,P} \cdot (g_{S,T} P) = g_{S,TP} g_{T,P} \qquad (S, T, P \text{ in } \mathfrak{G}). \tag{33}$$

Next put $S = T = I$ and $w_P = e$ and so obtain

$$(a \cdot x)P = aP \cdot xP, \tag{34}$$

that is, $\mathfrak{G}$ is a group of automorphisms of $\mathfrak{A}$. Put $x_T = w_P = e$, $P = I$, to see that the $g_{S,T}$ are in the center of $\mathfrak{A}$. Conversely if (33) holds for an automorphism group $\mathfrak{G}$ and the $g_{S,T}$ in the center of $\mathfrak{A}$ we have (32) and $\mathfrak{C}$ is associative.

We shall call an extension set satisfying (33) for an automorphism group $\mathfrak{G}$ of $\mathfrak{A}$ a *factor set*. Then we have proved

THEOREM 12. *A crossed extension $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ is associative if and only if $\mathfrak{A}$ is associative, $\mathfrak{G}$ is a group of automorphisms of $\mathfrak{A}$, $\mathfrak{g}$ is a factor set of $\mathfrak{A}$ whose quantities are in the center of $\mathfrak{A}$, and $\mathfrak{G} = \mathfrak{H}$.*

We now have the consequent

COROLLARY I. *Let $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$. Then if $\mathfrak{A}$ is not commutative the algebra $\mathfrak{C}$ is not associative.*

COROLLARY II. *Let $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$ where $\mathfrak{A}$ is a central simple algebra. Then $\mathfrak{C}$ is a non-associative central simple algebra.*

We have tacitly assumed in all of our work that the order n of $\mathfrak{A}$ is not one and that the order m of $\mathfrak{G}$ is also not one. Then $\mathfrak{C}$ has order nm and $\mathfrak{A}$ is a proper subalgebra of $\mathfrak{C}$.

10. Explicit construction

Let us indicate some of the types of algebras included under our definition. The first of these are the ordinary crossed products and the special case of cyclic algebras. These are given by $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ for $\mathfrak{G} = \mathfrak{H}$ the automorphism group of a normal field, $\mathfrak{g}$ a factor set. Then our construction Theorem 9 implies that $\mathfrak{C}$ is central simple and this seems to be a proof of that result which has been overlooked in the literature. The other algebras of Theorem 12 have been considered only in a rather special case.

Let us restrict all further attention to the case where $\mathfrak{A}$ is central simple and $\mathfrak{H}$ is the identity group so that $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$ is not associative and is central simple. Then we shall define one very interesting type of algebra which is the

crossed extension of $\mathfrak{A}$ by what is, essentially, a permutation group. We shall call such algebras *permutation algebras* and define them as follows. We let $e, u_2, \cdots, u_n$ be any basis of $\mathfrak{A}$ over $\mathfrak{F}$ and let u_1 be determined by

$$u_1 + \cdots + u_n = e. \tag{35}$$

Then $u_1, \cdots, u_n$ are a basis of $\mathfrak{A}$ over $\mathfrak{F}$ and we have a unique expression

$$a = \alpha_1 u_1 + \cdots + \alpha_n u_n \qquad (\alpha_i \text{ in } \mathfrak{F}) \tag{36}$$

for every a of $\mathfrak{A}$. We define

$$aS(P) = \alpha_1 u_{i_1} + \cdots + \alpha_n u_{i_n} \tag{37}$$

for every permutation

$$P = \begin{pmatrix} 1 & 2 \cdots n \\ i_1 & i_2 \cdots i_n \end{pmatrix} \tag{38}$$

and thus have defined a group $\mathfrak{G}$ of non-singular linear transformations $S = S(P)$ on $\mathfrak{A}$ such that $eS = e$ for every permutation group $\mathfrak{G}_0$ of permutations P. Clearly $\mathfrak{G}$ is equivalent to $\mathfrak{G}_0$ and the algebra $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{g})$ is central simple for every $\mathfrak{g}$. Moreover, this type of algebra is special since, while every finite group $\mathfrak{G}$ may be represented as a permutation group, $\mathfrak{G}$ may not[16] permute any set of basal quantities of A.

Let us give an iterative process next for the construction of a family of central simple algebras defined for what is essentially a single group. We let $\mathfrak{C} = (\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ be a given central simple crossed extension of order nm over $\mathfrak{F}$ defined for an algebra $\mathfrak{A}$ of order n and a group $\mathfrak{G}$ of order m. Then every quantity $\varkappa$ of $\mathfrak{C}$ is uniquely expressible in the form

$$\varkappa = u_1 \cdot a_1 + \cdots + u_m \cdot a_m$$

for the a_i in $\mathfrak{A}$ and $u_i = u_{S_i}$, $S_1 = I, S_2, \cdots, S_m$ the transformations of $\mathfrak{G}$. Since $\mathfrak{C}$ is central simple any

$$\mathfrak{C}_0 = (\mathfrak{C}, \mathfrak{G}_0, \mathfrak{g}_0)$$

will be central simple for any extending group $\mathfrak{G}_0$ and extension set $\mathfrak{g}_0$. We let $\mathfrak{G}_0$ be the set of linear transformations

$$S_0: \qquad \varkappa \to \varkappa S_0 = u_1 \cdot a_1 S + \cdots + u_m \cdot a_m S \qquad (S \text{ in } \mathfrak{G}).$$

Then $\mathfrak{G}_0$ is a finite group equivalent to $\mathfrak{G}$ and is clearly an extending group for $\mathfrak{C}$, the algebra $\mathfrak{C}_0 = (\mathfrak{C}, \mathfrak{G}_0, \mathfrak{g}_0)$ is central simple for every $\mathfrak{g}_0$ and we may

[16] It would be desirable now to prove the existence of examples of a simple algebra $\mathfrak{A}$ and an extending group $\mathfrak{G}$ such that no basis of $\mathfrak{A}$ exists for which $\mathfrak{G}$ may be regarded as a permutation group. Observe also that for every field $\mathfrak{A}$ of order n over $\mathfrak{F}$ we have a simple permutation algebra. This is then a generalization of the crossed product concept where $\mathfrak{A}$ is a normal field, the crossed product is a permutation algebra defined for a normal basis of $\mathfrak{A}$.

indeed choose $\mathfrak{g}_0$ in $\mathfrak{A}$. This process may be repeated to obtain central simple algebras of order nm^t for every $\mathfrak{G}$ of order m.

In particular we have the *generalized crossed products*

$$\mathfrak{C}_t = (\mathfrak{N}, \mathfrak{G}, \mathfrak{g}_1, \cdots, \mathfrak{g}_t),$$

where $\mathfrak{N}$ is a normal field of order r, $\mathfrak{G}$ is its automorphism group, the $\mathfrak{g}_i$ are all factor sets (or merely any extension sets). This algebra has order r^t over $\mathfrak{F}$. We thus have the *generalized cyclic algebras*

$$(\mathfrak{N}, S, \gamma_1, \cdots, \gamma_t)$$

for the $\gamma_i \neq 0$ in $\mathfrak{F}$.

Another process of iteration is that where we define S_0 by ${}_xS_0 = u_1 \cdot aS + u_2 \cdot a_2 + \cdots + u_m \cdot a_m$ and there are other obvious variations. However these may possibly give corresponding algebras obtained from the type given above by the use of a different $\mathfrak{G}$ and extension set. Thus we are led to the problem of determining when two crossed extensions defined for the same $\mathfrak{A}$ but with distinct groups and extension sets are equivalent, and also when they are isotopic. The solution of this problem will require a study of automorphisms of our algebras and also a solution of the simpler problem of determining conditions that $(\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ shall equal $(\mathfrak{A}, \mathfrak{G}, \mathfrak{H}, \mathfrak{k})$ for given distinct extension sets $\mathfrak{g}$ and $\mathfrak{k}$.

The associative algebra theory suggests for study many other fundamental problems regarding our new classes of algebras. For example let us call any algebra equivalent to a generalized cyclic algebra $(\mathfrak{N}, S, 1, \cdots, 1)$ a *generalized total matric algebra* over $\mathfrak{F}$. Then we seek to study the nature of the simple subalgebras of all such algebras (as well as of all other crossed extensions) and in particular to prove that they are all generalized cyclic algebras if $\mathfrak{F}$ is a p-adic or an algebraic number field. Such a study would probably require a study of splitting fields, direct products, the $\mathfrak{C}$-centralizer of a simple subalgebra of $\mathfrak{C}$, further extension of the concept of total matric algebra, similarity for crossed extensions, a theory of division algebras and a theory of exponents. It seems clear that our crossed extension definition includes many new varieties of simple algebras and it should lead to a host of new applications of modern algebraic techniques.

UNIVERSITY OF CHICAGO

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

EXTENSIONS OF DIFFERENTIAL FIELDS, I

BY E. R. KOLCHIN[1]

(Received May 6, 1942)

Introduction

It is a well-known theorem of algebra that a finite algebraic extension of a field of characteristic zero K always contains a primitive element ω:

$$K(\alpha_1, \cdots, \alpha_n) = K(\omega).$$

Moreover, by means of the theory of Galois, it is possible to characterize those elements of the extension which are primitive.[2] The present paper treats the analogous problems for differential fields (ordinary or partial).

A simple example shows that the precise analog is not true without further restriction. Let $\mathfrak{F}_0$ be the ordinary differential field of rational numbers, and let α_1 and α_2 be two algebraically independent complex constants. Since α_1 and α_2 both have zero derivatives, $\mathfrak{F}_0\langle\alpha_1, \alpha_2\rangle$ is set-theoretically identical with $\mathfrak{F}_0(\alpha_1, \alpha_2)$,[3] whence it is clear that there exists no number $\beta \in \mathfrak{F}_0\langle\alpha_1, \alpha_2\rangle$ such that $\mathfrak{F}_0\langle\alpha_1, \alpha_2\rangle = \mathfrak{F}_0\langle\beta\rangle$. However, for the theorem in question to hold it suffices to place a mild condition on the differential field. In the ordinary case the condition reduces to the requirement that *the differential field contain a non-constant* (that is, an element whose derivative is different from zero), in the general (partial) case, the condition is that *the differential field contain a set of elements whose Jacobian does not vanish.*

In studying those elements of an extension $\mathfrak{G}$ of a differential field $\mathfrak{F}$ which are primitive, a theorem presents itself which bears a similarity to results from Galois' theory. However any attempt in this direction seems destined to but fragmentary results, as the concept analogous to a *normal* extension of a field is lacking, so that one must speak of isomorphisms instead of automorphisms, thereby abandoning the concept of group.

1. Generic solutions

Throughout this paper $\mathfrak{F}$ will denote a differential field of characteristic zero with m types of differentiation $\delta_1, \cdots, \delta_m$,[4] and $y_1, \cdots, y_n$ will denote unknowns (m and n are positive integers).

Let Σ be a system of differential polynomials in $\mathfrak{F}\{y_1, \cdots, y_n\}$ with mani-

---

[1] National Research Fellow.

[2] This is not, of course, the simplest characterization.

[3] $\mathfrak{F}\langle u, \cdots\rangle$ means the result of the differential field adjunction to $\mathfrak{F}$ of the elements $u, \cdots$. $\mathfrak{F}(u, \cdots)$ means, as usual, the result of the field adjunction to $\mathfrak{F}$ (considered as a field) of the elements $u, \cdots$. The result of differential ring adjunction is indicated by curled brackets: $\mathfrak{F}\{u, \cdots\}$.

[4] This concept has been discussed by H. W. Raudenbush, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 714–720.

fold $\mathfrak{M}$. A set $\eta_1, \cdots, \eta_n$ of elements of a differential extension field of $\mathfrak{F}$ will be called *a generic solution of* Σ *(or of* $\mathfrak{M}$, *with respect to* $\mathfrak{F}$) if a necessary and sufficient condition for a differential polynomial $F(y_1, \cdots, y_n)$ in $\mathfrak{F}\{y_1, \cdots, y_n\}$ to belong to Σ is

$$F(\eta_1, \cdots, \eta_n) = 0.$$

It is easy to see that if Σ has a generic solution, then Σ is a prime differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$, so that $\mathfrak{M}$ is irreducible over $\mathfrak{F}$. Conversely if Σ is a prime differential ideal other than the whole ring $\mathfrak{F}\{y_1, \cdots, y_n\}$, then Σ has a generic solution. For example, if, in the differential ring of remainder classes $\mathfrak{F}\{y_1, \cdots, y_n\}/\Sigma$, $\bar{y}_i$ is the remainder class containing y_i, then $\bar{y}_1, \cdots, \bar{y}_n$ are elements of a differential field containing $\mathfrak{F}$ (namely, the differential field of quotients of $\mathfrak{F}\{y_1, \cdots, y_n\}/\Sigma$), and $F(y_1, \cdots, y_n)$ is in Σ if and only if $F(\bar{y}_1, \cdots, \bar{y}_n) = 0$. It is not hard to see moreover, that any generic solution $\eta_1, \cdots, \eta_n$ of Σ is equivalent to $\bar{y}_1, \cdots, \bar{y}_n$, that is, $\eta_i \to \bar{y}_i$ $(i = 1, \cdots, n)$ generates an isomorphism:

$$\mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle \cong \mathfrak{F}\langle\bar{y}_1, \cdots, \bar{y}_n\rangle.^5$$

Now, a prime differential ideal Σ in $\mathfrak{F}\{y_1, \cdots, y_n\}$ may very well decompose, over an extension $\mathfrak{G}$ of $\mathfrak{F}$, into several essential prime differential ideals:

$$\{\Sigma\} = \Lambda_1 \cap \cdots \cap \Lambda_s, \qquad \text{in } \mathfrak{G}\{y_1, \cdots, y_n\}. \tag{1}$$

Let $\zeta_1, \cdots, \zeta_n$ be a generic solution of some Λ_i, say of Λ_h. Then $\zeta_1, \cdots, \zeta_n$ *is a generic solution of* Σ. Indeed, it is clear that $F(y_1, \cdots, y_n) \in \Sigma$ implies $F(\zeta_1, \cdots, \zeta_n) = 0$, as $\Sigma \subseteq \Lambda_h$. Conversely, suppose that $F = F(y_1, \cdots, y_n) \in \mathfrak{F}\{y_1, \cdots, y_n\}$, and that $F(\zeta_1, \cdots, \zeta_n) = 0$. Let

$$G \in \Lambda_1 \cap \cdots \cap \Lambda_{h-1} \cap \Lambda_{h+1} \cap \cdots \cap \Lambda_s, \qquad G \notin \Lambda_h.$$

Then FG vanishes for all solutions of Σ, so that, by the Ritt analog of the *Nullstellensatz*, some power $(FG)^k$ is a linear combination, with coefficients in $\mathfrak{G}\{y_1, \cdots, y_n\}$, of differential polynomials in Σ:

$$F^kG^k = C_1S_1 + \cdots + C_lS_l \qquad (S_i \in \Sigma).$$

The coefficients of G^k, $C_1, \cdots, C_l$ are in $\mathfrak{G}$. Letting $\omega_1, \cdots, \omega_g$ be, with respect to $\mathfrak{F}$, a linearly independent linear basis of these coefficients, we find a relation

$$F^k(H_1\omega_1 + \cdots + H_g\omega_g) = T_1\omega_1 + \cdots + T_g\omega_g,$$

where each $T_i \in \Sigma$, each $H_i \in \mathfrak{F}\{y_1, \cdots, y_n\}$, and $H_1\omega_1 + \cdots + H_g\omega_g = G^k$. Equating coefficients, on both sides, of the linearly independent elements ω_i, we see that

[5] The isomorphism indicated by the symbol $\cong$ maps not only the sum and product of two elements onto the sum and product, respectively, of their images, but also the various derivatives of an element onto the corresponding derivatives of its image.

$$F^k H_i = T_i \in \Sigma \qquad (i = 1, \cdots, g).$$

But not every H_i is in Σ, for otherwise G would be in Λ_h. Hence, since Σ is a prime ideal, $F \in \Sigma$.

We use this result to prove that *if the prime differential ideal Σ in $\mathfrak{F}\{y_1, \cdots, y_n\}$ has a generic solution $\eta_1, \cdots, \eta_n$, if $\mathfrak{G}$ is a differential extension field of $\mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle$, and if no extension of $\mathfrak{G}$ contains another generic solution of Σ, then each $\eta_i \in \mathfrak{F}$.*

For, let (1) be the decomposition of $\{\Sigma\}$ into essential prime differential ideals in $\mathfrak{G}\{y_1, \cdots, y_n\}$. Any generic solution of Λ_1 is a generic solution of Σ and therefore is identical with $\eta_1, \cdots, \eta_n$. The same holds for every Λ_i, so that $s = 1$, $\{\Sigma\} = [y_1 - \eta_1, \cdots, y_n - \eta_n]$, and the only solution of Σ is $\eta_1, \cdots, \eta_n$. Assume, now, that $\eta_1 \notin \mathfrak{F}$. For some k,

$$(y_1 - \eta_1)^k = C_1S_1 + \cdots + C_lS_l \qquad (S_i \in \Sigma).$$

We suppose that k has been chosen as low as possible, so that $1, \eta_1, \cdots, \eta_1^k$ are linearly independent over $\mathfrak{F}$. Letting $1, \eta_1, \cdots, \eta_1^k, \omega_1, \cdots, \omega_q$ be a linearly independent linear basis, with respect to $\mathfrak{F}$, of the coefficients in $(y_1 - \eta_1)^k, C_1, \cdots, C_l$, and equating coefficients of η_1^k, we arrive at the contradiction that $1 \in \Sigma$. Hence $\eta_1 \in \mathfrak{F}$, similarly, every $\eta_i \in \mathfrak{F}$.

2. Relative isomorphisms

Let $\mathfrak{G}$ be a differential extension field of $\mathfrak{F}$. By an *isomorphism of* $\mathfrak{G}$ *with respect to* $\mathfrak{F}$ we shall mean an isomorphic mapping of $\mathfrak{G}$ onto a differential field $\mathfrak{G}'$ such that

(a) $\mathfrak{G}'$ is an extension of $\mathfrak{F}$,

(b) the isomorphic mapping leaves each element of $\mathfrak{F}$ invariant,

(c) $\mathfrak{G}$ and $\mathfrak{G}'$ have a common extension.

By means of well-ordering methods it is easy to show that an isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$ can be extended to an automorphism of the common extension of $\mathfrak{G}$ and its map under the isomorphism.

Concerning such relative isomorphisms we prove the following theorem:

Let $\mathfrak{G}$ be an extension of $\mathfrak{F}$, and let $\gamma \in \mathfrak{G}$. A necessary and sufficient condition that $\gamma \in \mathfrak{F}$ is that every isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$ leaves γ invariant. A necessary and sufficient condition that γ be a primitive element, that is, that $\mathfrak{G} = \mathfrak{F}\langle\gamma\rangle$, is that no isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$ other than the identity leaves γ invariant.

Proof: A. If $\gamma \in \mathfrak{F}$, then by condition (b), every isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$ leaves γ invariant. Now let $\gamma \notin \mathfrak{F}$, and denote by Γ the prime differential ideal of all differential polynomials in $\mathfrak{F}\{y\}$ which vanish for $y = \gamma$. γ is a generic solution of Γ. Since $\gamma \notin \mathfrak{F}$, we know by §1 that there exists a differential field $\mathfrak{H} \supseteq \mathfrak{G}$ in which Γ has another generic solution γ'. Now, $\gamma \to \gamma'$ generates an isomorphism between $\mathfrak{F}\langle\gamma\rangle$ and $\mathfrak{F}\langle\gamma'\rangle$ which leaves invariant every element of $\mathfrak{F}$. This isomorphism can be extended to an automorphism

of $\mathfrak{H}$, which automorphism in turn can be contracted to produce an isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$ which does not leave γ invariant.

B. If $\mathfrak{G} = \mathfrak{F}\langle\gamma\rangle$, every element of $\mathfrak{G}$ is a rational function, with coefficients in $\mathfrak{F}$, of γ and its various derivatives, so that an isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$ which leaves γ invariant leaves every element of $\mathfrak{G}$ invariant, that is, is the identity isomorphism. Conversely, if $\mathfrak{G} \neq \mathfrak{F}\langle\gamma\rangle$, there is an element $\alpha \in \mathfrak{G}$ such that $\alpha \notin \mathfrak{F}\langle\gamma\rangle$. By the part of the theorem already proved there is an isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}\langle\gamma\rangle$ which does not leave α invariant. This is an isomorphism of $\mathfrak{G}$ with respect to $\mathfrak{F}$, other than the identity, which leaves γ invariant.

The existence, in certain general cases, of a primitive element will be demonstrated in §4, after the proof of a preparatory result in §3.

3. Non-vanishing of nonzero differential polynomials

The following lemma will be used in §4.

A necessary and sufficient condition that, for an arbitrary nonzero differential polynomial $A = A(y_1, \cdots, y_n) \in \mathfrak{F}\{y_1, \cdots, y_n\}$, there exist elements $\eta_1, \cdots, \eta_n \in \mathfrak{F}$ such that $A(\eta_1, \cdots, \eta_n) \neq 0$, is that $\mathfrak{F}$ contain m elements $\xi_1, \cdots, \xi_m$ whose Jacobian is different from zero:

$$J = \begin{vmatrix} \delta_1 \xi_1 & \cdots & \delta_m \xi_1 \\ \cdot & \cdots & \cdot \\ \delta_1 \xi_m & \cdots & \delta_m \xi_m \end{vmatrix} \neq 0.$$

PROOF: Necessity. If $\mathfrak{F}$ has the property in question, then, in particular, there are elements $\xi_1, \cdots, \xi_m$ which do not annul

$$J(y_1, \cdots, y_m) = \begin{vmatrix} \delta_1 y_1 & \cdots & \delta_m y_1 \\ \cdot & \cdots & \cdot \\ \delta_1 y_m & \cdots & \delta_m y_m \end{vmatrix}.$$

Sufficiency. It obviously suffices to consider the case $n = 1$: $A = A(y) \in \mathfrak{F}\{y\}$. Now, since $J \neq 0$ there exists an $m \times m$ matrix (α_{ij}), with elements in $\mathfrak{F}$, such that $(\alpha_{ij})(\delta_j \xi_k)$ is the unit matrix. Hence, if we introduce the operators

$$\delta_i' = \alpha_{i1}\delta_1 + \cdots + \alpha_{im}\delta_m \qquad (i = 1, \cdots, m)$$

in terms of which, in turn, the operators δ_j may be expressed

$$\delta_j = \beta_{j1}\delta_1' + \cdots + \beta_{jm}\delta_m' \qquad (j = 1, \cdots, m),$$

we shall have

$$\delta_i'\xi_k = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

Moreover, since

$$\delta_p'\delta_q' = \sum_i \alpha_{pi}\delta_i \sum_j \alpha_{qj}\delta_j$$
$$= \sum_i \sum_j \alpha_{pi}\alpha_{qj}\delta_i\delta_j + \sum_j \left(\sum_i \alpha_{pi}\delta_i\alpha_{qj}\right)\delta_j,$$

we see that

$$\delta_p'\delta_q' = \delta_q'\delta_p' + \sum_k \gamma_k^{(p,q)}\delta_k' \qquad (\gamma_k^{(p,q)} \in \mathfrak{F}).$$

Hence $A(y)$ may be expressed as a polynomial, with coefficients in $\mathfrak{F}$, in the quantities $\delta_1'^{i_1} \cdots \delta_m'^{i_m} y$:

$$A(y) = P(\cdots, \delta_1'^{i_1} \cdots \delta_m'^{i_m} y, \cdots).$$

Letting the symbols $c_{i_1 \cdots i_m}$ denote constants in $\mathfrak{F}$ such that

$$P(\cdots, c_{i_1\cdots i_m}, \cdots) \neq 0,$$

and letting $\bar{a}_1, \cdots, \bar{a}_m$ be unknown constants (that is, indeterminates all of whose derivatives are zero), form the expression

$$\bar{\eta} = \sum \frac{c_{h_1 \cdots h_m}}{h_1! \cdots h_m!}(\xi_1 - \bar{a}_1)^{h_1} \cdots (\xi_m - \bar{a}_m)^{h_m}.$$

By the above, $\bar{\eta}$ satisfies the congruences

$$\delta_1'^{i_1} \cdots \delta_m'^{i_m}\bar{\eta} \equiv c_{i_1\cdots i_m} \qquad (\xi_1 - \bar{a}_1, \cdots, \xi_m - \bar{a}_m).$$

Hence

$$A(\bar{\eta}) \equiv P(\cdots, c_{i_1\cdots i_m}, \cdots) \qquad (\xi_1 - \bar{a}_1, \cdots, \xi_m - \bar{a}_m),$$

that is, $A(\bar{\eta})$ is a polynomial in the indeterminates $\bar{u}_j = \xi_j - \bar{a}_j$ $(j = 1, \cdots, m)$ with coefficients in $\mathfrak{F}$, and these coefficients are not all zero. Therefore we may choose rational values a_j for the unknown constants $\bar{a}_j$ so that, for

$$\eta = \sum \frac{c_{h_1 \cdots h_m}}{h_1! \cdots h_m!}(\xi_1 - a_1)^{h_1} \cdots (\xi_m - a_m)^{h_m},$$

we have $A(\eta) \neq 0$, q.e.d.

4. Existence of a primitive element

We are now in a position to prove our principal

THEOREM. *Let $\mathfrak{F}$ contain m elements whose Jacobian is different from zero. If $\mathfrak{F}\langle\alpha_1, \cdots, \alpha_n\rangle$ is a differential extension field of $\mathfrak{F}$ such that each α_i is a solution of a nonzero differential polynomial in $\mathfrak{F}\{y\}$, then there exists a primitive element γ:*

$$\mathfrak{F}\langle\alpha_1, \cdots, \alpha_n\rangle = \mathfrak{F}\langle\gamma\rangle.$$

By §2 we must show that there exists a $\gamma \in \mathfrak{F}\langle\alpha_1, \cdots, \alpha_n\rangle$ which is invariant

under no isomorphism of $\mathfrak{F}\langle\alpha_1, \cdots, \alpha_n\rangle$ with respect to $\mathfrak{F}$. We shall prove, as a lemma, a stronger result.

Let $A_i(y_i) \in \mathfrak{F}\{y_i\}$ have the solution $y_i = \alpha_i$ $(i = 1, \cdots, n)$. We shall show that *there exist elements* $\tau_1, \cdots, \tau_n \in \mathfrak{F}$ *such that* $\tau_1 y_1 + \cdots + \tau_n y_n$ *assumes different values for different solutions of* $\{A_1(y_1), \cdots, A_n(y_n)\}$.[6] Then certainly the element $\tau_1\alpha_1 + \cdots + \tau_n\alpha_n$ will satisfy our requirements on γ.

To prove this lemma, let $z_1, \cdots, z_n, t_1, \cdots, t_n$ be new unknowns, and, in $\mathfrak{F}\{y_1, \cdots, y_r, z_1, \cdots, z_n, t_1, \cdots, t_n\}$, consider the perfect differential ideal

$$\Omega = \{A_1(y_1), \cdots, A_n(y_n), A_1(z_1), \cdots, A_n(z_n), t_1(y_1 - z_1) + \cdots + t_n(y_n - z_n)\}.$$

Let $\Omega = \Omega_1 \cap \cdots \cap \Omega_s$ be the decomposition of Ω into essential prime differential ideals, and suppose the subscripts have been assigned so that $\Omega_1, \cdots, \Omega_r$ each contains every $y_i - z_i$, whereas $\Omega_{r+1}, \cdots, \Omega_s$ each fails to contain some $y_i - z_i$. Consider an Ω_j with $j > r$. Let $\bar{\eta}_1, \cdots, \bar{\eta}_n, \bar{\zeta}_1, \cdots, \bar{\zeta}_n, \bar{\tau}_1, \cdots, \bar{\tau}_n$ be a generic solution of Ω_j. Since $\bar{\tau}_1(\bar{\eta}_1 - \bar{\zeta}_1) + \cdots + \bar{\tau}_n(\bar{\eta}_n - \bar{\zeta}_n) = 0$, and some $\bar{\eta}_i - \bar{\zeta}_i$ is different from zero, $\bar{\tau}_1, \cdots, \bar{\tau}_n$ are dependent[7] over $\mathfrak{F}\langle\bar{\eta}_1, \cdots, \bar{\eta}_n, \bar{\zeta}_1, \cdots, \bar{\zeta}_n\rangle$. But each $\bar{\eta}_i$ and each $\bar{\zeta}_i$ annul a nonzero differential polynomial with coefficients in $\mathfrak{F}$. Hence $\bar{\tau}_1, \cdots, \bar{\tau}_n$ are dependent over $\mathfrak{F}$,[8] so that Ω_j contains a nonzero differential polynomial $L_j \in \mathfrak{F}\{t_1, \cdots, t_n\}$. Now let $M(t_1, \cdots, t_n) = L_{r+1} \cdots L_s$. By the authority of §3 choose elements $\tau_1, \cdots, \tau_n$ for which $M(\tau_1, \cdots, \tau_n) \neq 0$. For any two distinct solutions $y_i = \eta_i$ $(i = 1; \cdots, n)$ and $y_i = \zeta_i$ $(i = 1, \cdots, n)$ of $\{A_1(y_1), \cdots, A_n(y_n)\}$, the $3n$ elements

$$\eta_1, \cdots, \eta_n, \zeta_1, \cdots, \zeta_n, \tau_1, \cdots, \tau_n$$

cannot be a solution of Ω. For, these elements cannot be a solution of any Ω_j with $j \leqq r$ as each such Ω_j contains every $y_i - z_i$, and they cannot be a solution of an Ω_j with $j > r$ as each such Ω_j contains $M(t_1, \cdots, t_n)$. Consequently

$$\tau_1(\eta_1 - \zeta_1) + \cdots + \tau_n(\eta_n - \zeta_n) \neq 0.$$

Since $\eta_1, \cdots, \eta_n$ and $\zeta_1, \cdots, \zeta_n$ were chosen as *any* two distinct solutions of $\{A_1(y_1), \cdots, A_n(y_n)\}$, the proof of the lemma, and therefore of the theorem, is complete.

INSTITUTE FOR ADVANCED STUDY

---

[6] We lean heavily here on the proof for the ordinary case given by J. F. Ritt, *Differential equations from the algebraic standpoint*, American Mathematical Society Colloquium Publications, vol. XIV, New York, 1932. See especially pp. 26–31.

[7] See Raudenbush, loc. cit.

[8] See Raudenbush, loc. cit.

Annals of Mathematics
Vol. 43, No. 4, October, 1942

LE CORRESPONDANT TOPOLOGIQUE DE L'UNICITÉ DANS LA THÉORIE DES ÉQUATIONS DIFFÉRENTIELLES[1]

Par N. Aronszajn

(Received December 13, 1940; revised July 25, 1942)

Dans la théorie des équations différentielles, aussi bien ordinaires qu'aux dérivées partielles, on a pu établir des théorèmes d'existence et des théorèmes d'unicité. Il est apparu dans beaucoup de cas que, si pour les théorèmes d'existence il suffisait d'admettre pour les membres de l'équation des hypothèses de régularité très faibles, se réduisant parfois a la continuité seule (comme dans le cas de systèmes d'équations différentielles ordinaires), il était nécessaire d'admettre des hypothèses de régularité plus fortes pour assurer l'unicité.

La question se pose de caractériser dans les cas de multiplicité provenant de l'affaiblissement des hypothèses de régularité, l'ensemble des solutions multiples. Il apparait immédiatement que cette caractérisation doit tenir compte des propriétés topologiques de l'ensemble en question et que pour cela il est nécessaire d'introduire une topologie dans cet ensemble.

Sur cette voie nous sommes arrivé à établir une classe d'ensembles à laquelle appartiennent tous les ensembles des solutions multiples correspondant aux équations en question. Il nous semble probable que tout ensemble de cette classe est homéomorphe à l'ensemble des solutions multiples d'une équation du type considéré. Si cette suggestion était vraie, nous aurions eu ainsi une caractérisation topologique complète de ces ensembles de multiplicité et, en même temps, le correspondant topologique de l'unicité dans le cas de certaines types d'équations admettant de solutions multiples.

Les ensembles de la classe mentionnée seront désignés par R_δ. Ce sont des limites des suites décroissantes des ensembles R, ou par R nous désignons les retractes absolus de K. Borsuk.[2] Les R_δ conservent beaucoup de propriétés des retractes absolus.

Notre résultat principal peut être énoncé de manière intuitive (mais peu précise) comme suit: *Si les membres de l'équation en question peuvent être approchés aussi près que l'on veut par les membres d'une équation plus régulière, admettant une solution unique, l'ensemble des solutions de la première équation est un* R_δ.

Remarquons que, dans les cas particuliers que nous avons pu traiter, l'hypothèse de notre théorème concernant l'approximation avait pu être vérifiée grace au théorème de Weierstrass sur l'approximation d'une fonction continue par des

---

[1] Cet article forme un développement d'une conférence que l'auteur a faite le 19 avril à Paris, à une séance de la Société Math. de France. Les circonstances anormales actuelles n'ont pas permis de donner à cet article un développement aussi complet que l'aueur l'aurait souhaité. Surtout le côté bibliographique est en défau t, mais l'auteur n'a as pu faire mieux et il s'en excuse.

[2] Voir au sujet des rétractes les articles de K. Borsuk dans Fundamenta Math. a partir du t. 17 (1931) pp. 152-170.

polynomes, ou grace aux théorèmes similaires. A ce propos, rélevons que l'application de ce théorème de Weierstrass a déjà été faite par U. Müller[3] dans le cas de systèmes d'équations différentielles ordinaires, pour démontrer un théorème de H. Kneser. Ce dernier théorème, qui concerne le caractère continu de l'ensemble de solutions, est une simple conséquence de notre théorème (car R_δ est toujours un continu).

Le travail se compose de quatre paragraphes. Dans le §1 nous rappelons certains résultats et définitions essentiels. Dans le §2 nous prouvons un théorème auxiliaire concernant les suites de rétractes absolus. Le §3 est consacré au résultat fondamental de l'exposé. Des applications aux systèmes d'equations différentielles ordinaires forment le contenu du §4.

1. Résultats Préliminaires

D'après Borsuk[2] un rétracte absolu (R) est un espace métrique séparable qui est un rétracte de tout espace métrique qui le contient. Les rétractes absolus ont la propriété du point fixe, c'est-à-dire que toute représentation de R sur R possède un point invariant. Nous introduisons la notation R_δ pour désigner tout homéomorphe de l'intersection d'une suite décroissante de rétractes absolus. On peut aisément montrer que l'ensemble R_δ est un continuum à homologie et groupes fondamentaux ceux d'un point. Bien entendu, on sait que ces propriétés appartiennent aussi à R. Cependant R_δ et R peuvent différer en ce qui concerne leurs propriétés locales. Par example R_δ peut ne pas avoir de connexions locales, ainsi que le montre clairement l'exemple classique $y = \sin^2(\pi/x)$ pour $0 < x \leqq 1$ et $-1 \leqq y \leqq 1$ pour $x = 0$. Nous observons entre parenthèses qu'un R_δ dans le plan euclidien ne coupe pas le plan.

Dans le but de fournir des conclusions générales, nos resultats sont formulés pour certaines équations opérationelles de la forme

$$W = T(z)$$

dans les espaces de Banach.[4] Si T est continu et représente des ensembles bornés de E sur les ensembles (conditionellement) compactes de E', on dit alors que T est complètement continu. Notre contribution principale, le théorème C, est basée sur un théorème général d'existence du à Schauder.[5]

THÉORÈME A. *Lorsque T est complètement continu et représente K sur K, où K est borne, convexe et fermé dans E, son ensemble de points fixes est un sous-ensemble compacte en soi et non vide d'un R.*

L'equivalence avec la formulation de Schauder résulte du fait qu'un compactum convexe (ici l'extension convexe fermée[6] de $T(K)$), dans un espace de

[3] Voir M. Müller, Math. Zeitschrift, 28 (1928) pp. 619–645.

[4] S. Banach: *Théorie des Opérations Linéaires*, Warsaw 1932.

[5] Voir Math. Zeitschrift, 26 (1927) pp. 46–65 et Studia Math., 1 et 2. Des théorèmes de ce type ont déjà été donnés par G. D. Birkhoff et O. D. Kellogg, Transactions Amer. Math. Soc., 23 (1925) pp. 96–115; Lefschetz: *Topology* (New York 1930) p. 358 et Annals of Mathematics, vol. 38 (1937), pp. 819–822.

[6] S. Mazur: Studia, 2, (1930), pp. 7–10.

Banach est un R. Il serait intéressant de savoir si le théorème C, peut être étendu aux cas où les théorèmes d'existence (fondamentaux) sous-jacents sont démontrés par les méthodes, de Leray-Schauder.[7]

2. Les Suites de Rétractes Absolus

THÉORÈME B. *Soit* $\{R^{(n)}\}$ *une suite de rétractes absolus, sous-ensembles d'un même espace, et soit* M *un ensemble contenu dans tous les* $R^{(n)}$. *Si les* $R^{(n)}$ *convergent vers* M, *ce dernier ensemble est un* R_δ.

Démonstration. Soit $\mathfrak{S}$ l'espace contenant tous les $R^{(n)}$ et soit φ_n une fonction rétractant $\mathfrak{S}$ sur $R^{(n)}$.

Nous pouvons toujours supposer que l'espace $\mathfrak{S}$ est *distanciable* et que l'on a choisi pour lui une distance $\rho(x, y)$ bornée supérieurement (autrement nous aurions pu remplacer $\mathfrak{S}$ par la somme de tous les $R^{(n)}$ qui a certainement ces propriétés).

Nous allons choisir une sous-suite $\{\bar{R}^{(k)}\}$ de $\{R^{(n)}\}$ de sorte que, en désignant par $\bar{\varphi}_k$ la fonction φ_n correspondant a $\bar{R}^{(k)}$, et par $\psi_i^{(k)}$, pour $i < k$, la fonction composée

$$\psi_i^{(k)} = \bar{\varphi}_i\bar{\varphi}_{i+1} \cdots \bar{\varphi}_{k-1}, \tag{1}$$

on ait pour tous k, $i < k$ et $x \in \bar{R}^{(k)}$,

$$\rho(x, \psi_i^{(k)}(x)) \leqq 1/k. \tag{2}$$

Pour définir les $\bar{R}^{(k)}$ nous commençons par poser $\bar{R}^{(1)} = R^{(1)}$. Supposons maintenant que les $\bar{R}^{(1)}, \bar{R}^{(2)}, \cdots, \bar{R}^{(k)}$ sont déjà définis. D'après (1), $\psi_i^{(k+1)}$ est alors défini, et on a pour tout x de M et tout $i < k + 1$

$$x = \psi_i^{(k+1)}(x),$$

car pour tout r, $M \subset \bar{R}^{(r)}$, et par conséquent $\psi_r(x) = x$. Il s'ensuit qu'il existe un voisinage V de M tel que, pour $x \in V$ et tout $i < k + 1$, on ait,

$$\varphi(x, \psi_i^{(k+1)}(x)) \leqq \frac{1}{k+1}.$$

Nous poserons $\bar{R}^{(k+1)}$ = le premier $R^{(n)}$ postérieur à $\bar{R}^{(k)}$ dans la suite $\{R^{(n)}\}$, contenu dans V. Il est clair qu'un tel $R^{(n)}$ existe vu que les $R^{(n)}$ convergent vers M. Ainsi, les $R^{(k)}$ se définissent successivement et la propriété (2) est remplie.

Considérons maintenant le produit combinatoire infini $\mathfrak{S}^\infty = \mathfrak{S} \times \mathfrak{S} \times \cdots$ aux elements $X = (x_1, x_2, \cdots, x_n, \cdots)$ avec $x_n \in \mathfrak{S}$. On définit dans $\mathfrak{S}^\infty$ une distance à la Fréchet

$$\rho(X, Y) = \sum_{n=1}^{\infty} 2^{-n} \rho(x_n, y_n).$$

---

[7] Voir J. Leray et J. Schauder, Annales Scient. École Norm. Sup., 51 (1934) pp. 45–78.

La notion de limite correspondante se définit comme suit: la suite $\{X^{(k)}\}$ converge vers X, si chaque suite $\{x_n^{(k)}\}$ converge vers x_n.

Considerons dans $\mathfrak{E}^\infty$ les sous-ensembles $Q^{(k)}$, définis de manière suivante: $Q^{(k)}$ est composé de tous les points $X = (x_1, x_2, \cdots, x_k, \cdots)$ tels que, pour $n \geqq k$, $x_n \in \bar{R}^{(n)}$, tandis que pour $n < k$, $x_n = \psi_n^{(k)}(x_k)$.

Il est clair que $Q^{(k)}$ est homéomorphe avec l'ensemble de toutes les suites $(x_k, x_{k+1}, \cdots)$ où x_n parcourt $\bar{R}^{(n)}$, $n = k, k+1, \cdots$. Cet ensemble forme le produit combinatoire $R^{(k)} \times R^{(k+1)} \times \cdots$ de rétractes absolus $R^{(n)}$; c'est donc un rétracte absolu.[8] Il en résulte que $Q^{(k)}$ *est un rétracte absolu.*

Remarquons ensuite que $Q^{(k)} \supset Q^{(k+1)}$. En effet, si $X = (x_1, x_2, \cdots, x_k, x_{k+1}, \cdots)$ appartient à $Q^{(k+1)}$, on a d'après la définition de $Q^{(k+1)}$: $x_n \in \bar{R}^{(n)}$ pour $n \geqq k+1$, $x_k = \psi_k^{(k+1)}(x_{k+1}) = \bar{\varphi}_k(x_{k+1}) \in \bar{R}^{(k)}$ et enfin, pour $n < k$, $x_n = \psi_n^{(k+1)}(x_{k+1}) = \bar{\varphi}_n\bar{\varphi}_{n+1} \cdots \bar{\varphi}_k(x_{k+1}) = \psi_n^{(k)}\bar{\varphi}_k(x_{k+1}) = \psi_n^{(k)}(x_k)$, donc $X \in Q^{(k)}$.

Prouvons maintenant que la suite décroissante $Q^{(1)}, Q^{(2)}, \cdots$ a pour intersection l'ensemble M' composé de tous les $X = (x_1, x_2, \cdots)$ avec $x_1 = x_2 = x_3 = \cdots = x \in M$. En effet, si $X = (x_1, x_2, \cdots)$ appartient a tous les $Q^{(k)}$, on aura suivant (2) pour tout k et tout $n < k$

$$\rho(x_k, x_n) = \rho(x_k, \psi_n^{(k)}(x_k)) \leqq \frac{1}{k}.$$

Il en résulte que tout x_n, $n = 1, 2, \cdots$, est la limite de la suite $\{x_k\}$ qui est nécessairement convergente. Il s'ensuit d'une part que $x_1 = x_2 = \cdots = x$. D'autre part, $x_k \in \bar{R}^{(k)}$ et, les $R^{(k)}$ convergeant vers M, la limite x de $\{x_k\}$ appartient à M. Ainsi $M' \supset Q^{(1)}Q^{(2)} \cdots$. Inversement, si $X \in M'$, il appartient à tout $Q^{(k)}$, car, pour $n \geqq k$, $x_n = x \in M \subset \bar{R}^{(n)}$ et, pour $n < k$, $x_n = x = \psi_n^{(k)}(x) = \psi_n^{(k)}(x_k)$, vu que toute φ_i transforme un $x \in M$ en lui-même. Il est donc prouve que $M' = Q^{(1)}Q^{(2)} \cdots$.

Enfin, il est évident que l'ensemble M' est homéomorphe avec M par l'intermédiaire de la correspondance donnant à un $X = (x, x, x \cdots)$ de M', pour image le point x de M.

Ainsi, M est homéomorphe avec M' qui est, d'après ce qui précède, un R_δ. M est donc lui-même aussi un R_δ, c.q.f.d.

3. Le théorème principal

Pour pouvoir poursuivre nos raisonnements nous allons admettre que *la transformation T peut être approchée aussi près que l'on veut par une transformation "plus régulière."*

Pour préciser cette hypothèse revenons aux notations du §1 précédent. Nous supposerons qu'à tout $\epsilon > 0$ on peut faire correspondre une transformation T_ϵ complètement continue de l'espace E en lui-même de sorte que

1°. $\| T_\epsilon(z) - T(z) \| \leqq \epsilon$ *pour tout élément z de K*, $\| \ \|$ désignant la norme dans E;

[8] N. Aronszajn et K. Borsuk, Fundamenta, 18, 1932, pp. 193–197.

2°. *La transformation* $z_1 = z - T_\epsilon(z) \equiv H_\epsilon(z)$ *représente de manière biunivoque l'ensemble* K *en un ensemble contenant une sphère* $\|z\| \leqq \rho$, *avec* ρ *indépendant de* ϵ.

Tandis que la première condition précise de manière dont T est approchée par T_ϵ, la seconde condition peut être applée "condition d'unicité," car elle a pour conséquence l'existence et l'unicité (si l'on se limite aux solutions appartenant à K) de la solution de l'équation en z

$$z - T_\epsilon(z) = z_1,$$

pour z_1 de norme suffisamment petite. Comme nous l'avons déjà remarqué, pour avoir l'unicité de solution il faut admettre en général des conditions supplémentaires de régularité, c'est pourquoi nous dirons que T_ϵ est "plus régulière" que T.

Dans ces conditions nous pouvons démontrer le

THÉORÈME C. *L'ensemble des solutions de l'équation* $T(z) = z$ *est un* R_δ.

DÉMONSTRATION. Désignons cet ensemble des solutions par S. Considérons les transformations $T_n \equiv T_{\epsilon_n}$ pour une suite $\{\epsilon_n\}$ tendant vers 0, tous les ϵ_n étant $\leqq \rho$.

Considérons d'abord les transformations T_n et H_n pour un n fixe. L'ensemble S étant contenu dans K, on a d'après 1°, pour tout élément ζ de S,

$$\|H_n(\zeta)\| = \|H_{\epsilon_n}(\zeta)\| = \|\zeta - T_n(\zeta)\| = \|T(\zeta) - T_n(\zeta)\| \leqq \epsilon_n.$$

Par conséquent, l'ensemble transformé $H_n(S)$ est contenu dans la sphère de rayon $\epsilon_n \leqq \rho$. Cet ensemble, obtenu par transformation continue d'un ensemble compact en soi, est compact en soi (voir le théorème A). Le plus petit corps convexe le contenant est aussi compact en soi et compris dans la sphère de rayon $\epsilon_n \leqq \rho$.

Soit Q_n ce corps convexe. D'après la condition 2°, la transformation H_n^{-1} inverse de la transformation H_n, est définie sur Q_n. Elle donne de Q_n une image $R^{(n)} = H_n^{-1}(Q_n)$ contenue dans l'ensemble K.

La transformation inverse H_n^{-1} n'est pas en général continue, mais nous allons montrer qu'elle l'est sur Q_n. En effet, si une suite d'éléments $\{h_k\}$ de Q_n tend vers h (qui appartient aussi à Q_n, celui-ci étant compact en soi), on a pour les éléments $z_k = H_n^{-1}(h_k)$ les équations suivantes

$$z_k - T_n(z_k) = h_k.$$

Les z_k appartenant a l'ensemble borné K, ils forment une suite bornée, et la transformation complètement continue T_n transforme cette suite en une suite compacte. Si les z_k ne tendaient pas vers l'élément $z = H_n^{-1}(h)$ donné par l'équation

$$z - T_n(z) = h,$$

on pourrait extraire des z_k une suite $\{z_{k_i}\}$ n'admettant pas z comme élément limite et telle que les $T_n(z_{k_i})$ convergent vers un élément g. Mais alors les

éléments $z_{k_i} = T_n(z_{k_i}) + h_{k_i}$ convergeraient vers $g + h$ et les $T_n(z_{k_i})$ convergeraient vers $T_n(g + h)$, qui serait égal à g, et on aurait

$$g + h = T_n(g + h) + h;$$

$g + h$ serait donc la solution z de $z - T_n(z) = h$ et les z_{k_i} y convergeraient, d'où contradiction.

Cette contradiction prouve que, sur Q_n, la transformation H_n^{-1} est continue. Puisque son inverse H_n est d'après 2° biunivoque et continue, la transformation H_n^{-1} représente de manière *homéomorphe* Q_n en $R^{(n)} = H_n^{-1}(Q_n)$.

L'ensemble Q_n étant compact en soi et convexe, c'est un *rétracte absolu*. Par conséquent, son homéomorphe $R^{(n)}$ l'est aussi. D'autre part Q_n contenait le transformé $H_n(S)$ de S. Il s'ensuit que $R^{(n)} = H_n^{-1}(Q_n)$ contient S. Dès lors, pour prouver notre théorème, il nous reste a prouver que les $R^{(n)}$ convergent vers S pour n tendant vers l'∞.

A cet effet prenons une suite quelconque $\{z_j\}$ telle que chaque z_j appartient à un $R^{(n_j)}$, les n_j tendant vers l'∞. Comme nous l'avons vu plus haut, tous les $R^{(n)}$ sont contenus dans l'ensemble borné K. Il s'ensuit que $\{z_j\}$ est une suite bornée et que les $T(z_j)$ forment une suite compacte de laquelle on peut extraire une sous-suite $\{T(z_{j_k})\}$ convergeant vérs un élément z. Les z_j appartenant à K, on a selon 1°

$$|| T_{n_{j_k}}(z_{j_k}) - T(z_{j_k}) || \leqq \epsilon_{n_{j_k}} \to 0,$$

donc, les $T_{n_{j_k}}(z_{j_k})$ convergent aussi vers z. D'après la définition des $R^{(n)}$, on a pour tout z_{j_k} l'équation

$$z_{j_k} = T_{n_{j_k}}(z_j) + h_{j_k},$$

où h_{j_k} appartient à $Q_{n_{j_k}}$, donc à une sphère de rayon $\epsilon_{n_{j_k}} \to 0$. Il en résulte successivement: $\lim z_{j_k} = \lim T_{n_{j_k}}(z_{j_k}) = z$, $\lim T(z_{j_k}) = T(z)$ donc $T(z) = z$. Ainsi, de toute suite $\{z_j\}$ avec z_j appartenant à $R^{(n_j)}$ on peut extraire une suite $\{z_{j_k}\}$ convergeant vers une solution z de l'equation $T(z) = z$, donc vers un élément de S. Ceci prouve que les $R^{(n)}$ convergent vers S.

Notre théorème est ainsi démontré.

4. Application

Comme application de notre théorème général nous allons considérer un système d'équations différentielles ordinaires. Sans restreindre essentiellement la généralité nous pouvons nous limiter au cas d'un système de deux équations avec deux fonctions inconnues

$$(3) \qquad \frac{dx}{dt} = u(x, y, t), \qquad \frac{dy}{dt} = v(x, y, t).$$

avec les conditions initiales

$$x(0) = 0, \qquad y(0) = 0.$$

Il est connu depuis Peano que ce système admet certainement de solutions, si seulement u et v sont continues. En général ces solutions existeront dans un intervalle de t entourant $t = 0$. Si nous supposons—ce que nous allons faire dans la suite—que les fonctions u et v sont continues et bornées pour toutes les valeurs de x, y et t, *les solutions existeront sur tout l'axe de* t.

D'autre part, on sait que si l'on suppose les fonctions u et v satisfaisant à une condition de Lipschitz relativement à x et y, uniformément en t dans tout intervalle fini, *la solution est unique*. Elle le sera donc à fortiori, si u et v sont analytiques en x, y et t, ou ne diffèrent d'une telle fonction que par une fonction de t seul.

Pour appliquer notre théorème général, neus envisagerons l'espace vectoriel de tous les couples de fonctions $[x(t), y(t)]$ admettant des dérivées $x'(t)$ et $y'(t)$ continues, et satisfaisant aux conditions $x(0) = y(0) = 0$. Nous considérerons ces fonctions dans un intervalle fini fixe $\alpha \leqq t \leqq \beta$, $\alpha < 0 < \beta$, arbitrairement choisi.

Dans cet espace vectoriel nous prendrons comme norme d'un couple de fonctions

$$z = [x(t), y(t)],$$

le nombre

$$\| z \| = \max_{\alpha \leqq t \leqq \beta} [x'(t)^2 + y'(t)^2]^{\frac{1}{2}}.$$

Considérons dans cet espace la transformation

$$z_1 = T(z) = [x_1(t), y_1(t)], \qquad x_1(t) = \int_0^t u(x, y, t)\, dt, \, y_1(t) = \int_0^t v\, dt,$$

où $z = [x(t), y(t)]$.

Il est clair que, si l'on pose

$$m = \text{borne sup } (u^2 + v^2)^{\frac{1}{2}}, \textit{ pour tous les } x, y, t,$$

la sphère de notre espace vectoriel,

$$\| z \| \leqq 2m$$

peut être prise comme l'ensemble K dans la théorie générale, car la transformation T la représente en elle-même. D'autre part, on prouve facilement que T est complètement continue. Ceci permet déjà d' appliquer le théorème d'existence A.

Pour appliquer le théorème C, il faut définir les transformations T_ϵ conformément aux conditions 1° et 2° du §3. A cet effet remarquons d'abord que, si

pour $z = [x(t), y(t)]$ on a $\| z \| \leqq 2m$, il en résulte pour $x(t)$ et $y(t)$, $\alpha \leqq t \leqq \beta$, les inégalités

$$| x(t) | \leqq 2m(\beta - \alpha), \qquad | y(t) | \leqq 2m(\beta - \alpha).$$

Par conséquent, pour satisfaire aux conditions 1° et 2°, il suffira d'approcher chacune des fonctions $u(x, y, t)$ et $v(x, y, t)$ par des fonctions analytiques u_ϵ et v_ϵ des trois variables reeles x, y et t, satisfaisant aux inégalités

$$|u_\epsilon - u| \leqq \frac{\epsilon}{2}, \qquad | v_\epsilon - v | \leqq \frac{\epsilon}{2}, \textit{ pour}$$

$$| x | \leqq 2m(\beta - \alpha), \quad | y | \leqq 2m(\beta - \alpha), \quad \alpha \leqq t \leqq \beta,$$

$$| u_\epsilon | \leqq m\sqrt{2}, \ | v_\epsilon | \leqq m\sqrt{2} \textit{ pour tous les } x, y, t.$$

En se basant sur le théorème d'approximation de Weierstrass on construit aisément les fonctions u_ϵ et v_ϵ. Les théorèmes d'existence et d'unicité, indiqués au commencement de ce § permettent de vérifier immédiatement la condition 2°. Ainsi, le théorème C est applicable.

Indiquons quelques conséquences de ce théorème dans le cas présent. Comme on sait, à chaque solution de notre système avec les conditions initiales $x(0) = y(0) = 0$, correspond dans l'espace des variables x, y, t une courbe intégrale du système passant par l'origine $x = y = t = 0$. S'il y a plusieurs solutions, il passe par l'origine tout un faisceau de courbes intégrales. Chacune de ces courbes coupe le plan $t = t_0$ au point $x_0 = x(t_0)$, $y_0 = y(t_0)$ qui varie de façon continue quand la courbe intégrale parcourt le faisceau. Il s'ensuit que la trace du faisceau sur le plan $t = t_0$ est une image continue du faisceau, c'est à dire de l'ensemble S des solutions du système (3). Cet ensemble étant un R_δ, donc à fortiori un continu, son image est également un continu. Par consequent, *la trace sur le plan $t = t_0$ du faisceau des courbes intégrales passant par l'origine —ou par un point quelconque—est un continu.* C'est le théorème de Kneser; il se montre ainsi une conséquence immédiate de notre théorème.

Dans des cas particuliers nous pouvons préciser la nature de cette trace. Par exemple, si les fonctions $u(x, y, t)$ et $v(x, y, t)$ satisfont dans tout l'espace des x, y, t a la même condition de Lipschitz sauf au point $x = y = t = 0$, il n'y aura dans tout l'espace que l'origine comme point par lequel puissent passer plusieurs courbes intégrales. Dans ce cas, chaque point de la trace sur le plan $t = t_0 \neq 0$ ne provient que d'une seule courbe intégrale. Par conséquent, *la trace est une image homéomorphe de l'ensemble S et est un R_δ.* D'après la propriété caractéristique des R_δ plans, cette trace ne coupe pas le plan $t = t_0$.

Il est très probable que tout R_δ plan peut être obtenu comme trace du faisceau intégral pour un choix convenable des fonctions u et v conformes aux conditions ci-dessus. Ceci est en rapport avec l'hypothèse que nous avons émise dans l'introduction.

Il serait intéressant d'étudier la nature du faisceau intégral et de ses traces pour différentes classes de fonctions. En particulier, on pourrait étudier la nature topologique du faisceau intégral pour les fonctions u et v continues et bornées dans tout l'espace (x, y, t) et analytiques partout sauf à l'origine.

EDITORS NOTE. Owing to present circumstances it was impossible to communicate freely with the author regarding certain necessary revisions in the paper. With the authorization of the author and some information conveyed by him, this was accomplished by Professor D. G. Bourgin, to whom the Editors wish to express their personal thanks and also those of the author. S. L.

LONDON

ANNALS OF MATHEMATICS
Vol. 43, No. 4, October, 1942

A PROOF THAT THERE EXISTS A CIRCUMSCRIBING CUBE AROUND ANY BOUNDED CLOSED CONVEX SET IN R^3

BY SHIZUO KAKUTANI

(Received June 8, 1942)

1

The following problem was proposed by Professor Rademacher: Given a bounded closed convex set in a three-space R^3, is it always possible to find a circumscribing cube around it? It is easy to see (cf. §3) that this problem can be reduced to the following one: Given a real-valued continuous function $f(P)$ defined on a two-sphere S^2, is it possible to find a triple of points $P_1, P_2, P_3 \in S^2$, perpendicular to one another (this means that the three vectors $\mathbf{OP}_1$, $\mathbf{OP}_2$, $\mathbf{OP}_3$ from the center O of S^2 to these three points P_1, P_2, P_3 are perpendicular to one another) such that $f(P_1) = f(P_2) = f(P_3)$? The purpose of the present note is to answer these questions in the affirmative.

2

THEOREM 1. *Let $f(P)$ be a real-valued continuous function defined on a two-sphere S^2. Then there exists a triple of points $P_1, P_2, P_3 \in S^2$, perpendicular to one another, such that $f(P_1) = f(P_2) = f(P_3)$.*

PROOF. Let us consider S^2 as a sphere of radius 1 in a three-space R^3, with the origin $O = (0, 0, 0)$ of R^3 as a center. Let us put $P_1^\circ = (1, 0, 0)$, $P_2^\circ = (0, 1, 0)$, $P_3^\circ = (0, 0, 1)$. Let further $G = \{\sigma\}$ be the group of all rotations of S^2 (or equivalently, rotations of R^3 around its origin $O = (0, 0, 0)$). G is a three dimensional compact manifold.

For any $\sigma \in G$, consider the point $\varphi(\sigma) = (x, y, z) \in R^3$ defined by $x = f(\sigma^{-1}(P_1^\circ))$, $y = f(\sigma^{-1}(P_2^\circ))$, $z = f(\sigma^{-1}(P_3^\circ))$. It is clear that $\sigma \to \varphi(\sigma)$ is a continuous mapping of G into R^3. In order to prove our theorem, it suffices to show that there exists a rotation $\sigma \in G$ such that $\varphi(\sigma)$ lies on the straight line $l: x = y = z$ in R^3.

We assume the contrary, and shall draw a contradiction from it. Let ρ be the projection of R^3 onto the plane $\pi: x + y + z = 0$, which is perpendicular to the line l. Then $\sigma \to \psi(\sigma) \equiv \rho(\varphi(\sigma))$ is a continuous mapping of G into π. By assumption, the image $\psi(G)$ of G by this mapping $\psi(\sigma)$ does not contain the origin $O = (0, 0, 0)$.

Let H be the subgroup of G consisting of all rotations around the line l. H is isomorphic to the group of rotations of the plane π around the origin $O = (0, 0, 0)$, and we may denote elements of H by $\sigma_\theta (0 \leqq \theta \leqq 2\pi)$, where θ denotes the angle of rotation around the axis l, measured in such a sense that we have $\sigma_{2\pi/3}(P_i^\circ) = P_{i+1}^\circ$, $i = 1, 2, 3, \bmod 3$.

Let us denote the rotation of the plane π around its origin $O = (0, 0, 0)$, which corresponds to σ_θ, by $\tau_\theta (0 \leqq \theta \leqq 2\pi)$. It is then easy to see that we have

$$\psi(\sigma_{\theta+2(\pi/3)}) = \tau_{2\pi/3}(\psi(\sigma_\theta)), \qquad \psi(\sigma_{\theta+(4\pi/3)}) = \tau_{4\pi/3}(\psi(\sigma_\theta))$$

for any $\theta (0 \leqq \theta \leqq 2\pi)$. Let C_{θ_1, θ_2} be the curve traced on π by $\psi(\sigma_\theta)$ when θ runs over the interval $\theta_1 \leqq \theta \leqq \theta_2$. Then the fact stated above means that the curves $C_{2\pi/3, 4\pi/3}$ and $C_{4\pi/3, 2\pi}$ are obtained by applying the rotations $\tau_{2\pi/3}$ and $\tau_{4\pi/3}$ to the curve $C_{0, 2\pi/3}$.

Let α be the increment of the angle around the origin $O = (0, 0, 0)$ in the plane π, when the point $\psi(\sigma_\theta)$ runs over the curve $C_{0, 2\pi/3}$ from $\psi(\sigma_0)$ to $(\sigma_{2\pi/3})$, or equivalently, when θ runs from 0 to $2\pi/3$. Then α must be of the form: $\alpha = 2m\pi + 2\pi/3$, where m is an integer ($m = 0, \pm 1, \pm 2, \cdots$). Hence as θ runs from 0 to 2π, the total increment of the angle of $\psi(\sigma_\theta)$ around the origin $O = (0, 0, 0)$ in the plane π is $6m\pi + 2\pi = (3m + 1)\cdot 2\pi$.

On the other hand, consider H as a closed curve on the manifold of the topological group G. Then it is well known that $2H$ is homotopic to zero on G. Consequently, the curve $2C_{0, 2\pi}$, which is the image of $2H$ by the mapping $\sigma \to \psi(\sigma)$, must also be homotopic to zero on π^*, where by π^* we mean the open set which is obtained by taking away the origin $O = (0, 0, 0)$ from the plane π. This is, however, impossible, since the total increment of the angle on the curve $2C_{0, 2\pi}$ is $2(3m + 1)\cdot 2\pi \neq 0$.

Thus we arrive at a contradiction, and the proof of Theorem 1 is completed.

3

THEOREM 2. *Let K be a bounded closed convex set in a three space R^3. Then there exists a circumscribing cube around K.*

PROOF. Let S^2 be a two sphere in R^3 with the origin $O = (0, 0, 0)$ of R^3 as a center. For any point $P \in S^2$, consider two tangent planes to K (parallel to each other) which are perpendicular to the vector **OP**. These two planes may coincide if K is a flat convex set. Let $f(P)$ be the vertical distance of these two planes. $f(P)$ is clearly a real-valued continuous function defined on S^2. (Moreover, $f(P)$ takes the same value at two antipodal points of S^2; but we do not need this fact in our proof). By Theorem 1, there exists a triple of points $P_1, P_2, P_3 \in S^2$, perpendicular to one another, such that $f(P_1) = f(P_2) = f(P_3)$. It is then clear that the corresponding six tangent planes form a cube which is circumscribing around the convex set K.

4. Remarks

There are two problems related to our results. The first one is to investigate whether it is possible to inscribe a cube in a given bounded open convex set in R^3. The answer to this question is negative, and a counter-example to this is given by a tetrahedron in R^3 which is extremely flat. In fact, if we take a convex quadrangle $ABCD$ on the (x, y)-plane, such that the two diagonals AC and BD are not perpendicular to each other, and if we shift the vertex A in a direction of z-axis by a small distance, then the tetrahedron $A'BCD$ thus obtained is a required one. It is easy to see that there is no inscribing cube in this tetrahedron.

The second problem concerns the possibility of generalizations to higher dimensional cases. It is not yet known whether or not it is possible to find a circumscribing n-dimensional cube around any given bounded closed convex set in R^n ($n \geqq 4$). We may also ask: Given a real-valued continuous function $f(P)$ defined on an $(n - 1)$-sphere S^{n-1}, is it possible to find n points $P_1, \cdots, P_n$ on S^{n-1}, perpendicular to one another, such that $f(P_1) = \cdots = f(P_n)$ ($n \geqq 4$)? These problems are still unsolved.

INSTITUTE FOR ADVANCED STUDY

for any $\theta (0 \leqq \theta \leqq 2\pi)$. Let C_{θ_1, θ_2} be the curve traced on π by $\psi(\sigma_\theta)$ when θ runs over the interval $\theta_1 \leqq \theta \leqq \theta_2$. Then the fact stated above means that the curves $C_{2\pi/3, 4\pi/3}$ and $C_{4\pi/3, 2\pi}$ are obtained by applying the rotations $\tau_{2\pi/3}$ and $\tau_{4\pi/3}$ to the curve $C_{0, 2\pi/3}$.

Let α be the increment of the angle around the origin $O = (0, 0, 0)$ in the plane π, when the point $\psi(\sigma_\theta)$ runs over the curve $C_{0, 2\pi/3}$ from $\psi(\sigma_0)$ to $(\sigma_{2\pi/3})$, or equivalently, when θ runs from 0 to $2\pi/3$. Then α must be of the form: $\alpha = 2m\pi + 2\pi/3$, where m is an integer ($m = 0, \pm 1, \pm 2, \cdots$). Hence as θ runs from 0 to 2π, the total increment of the angle of $\psi(\sigma_\theta)$ around the origin $O = (0, 0, 0)$ in the plane π is $6m\pi + 2\pi = (3m + 1)\cdot 2\pi$.

On the other hand, consider H as a closed curve on the manifold of the topological group G. Then it is well known that $2H$ is homotopic to zero on G. Consequently, the curve $2C_{0, 2\pi}$, which is the image of $2H$ by the mapping $\sigma \to \psi(\sigma)$, must also be homotopic to zero on π^*, where by π^* we mean the open set which is obtained by taking away the origin $O = (0, 0, 0)$ from the plane π. This is, however, impossible, since the total increment of the angle on the curve $2C_{0, 2\pi}$ is $2(3m + 1)\cdot 2\pi \neq 0$.

Thus we arrive at a contradiction, and the proof of Theorem 1 is completed.

3

THEOREM 2. *Let K be a bounded closed convex set in a three space R^3. Then there exists a circumscribing cube around K.*

PROOF. Let S^2 be a two sphere in R^3 with the origin $O = (0, 0, 0)$ of R^3 as a center. For any point $P \epsilon S^2$, consider two tangent planes to K (parallel to each other) which are perpendicular to the vector **OP**. These two planes may coincide if K is a flat convex set. Let $f(P)$ be the vertical distance of these two planes. $f(P)$ is clearly a real-valued continuous function defined on S^2. (Moreover, $f(P)$ takes the same value at two antipodal points of S^2; but we do not need this fact in our proof). By Theorem 1, there exists a triple of points $P_1, P_2, P_3 \epsilon S^2$, perpendicular to one another, such that $f(P_1) = f(P_2) = f(P_3)$. It is then clear that the corresponding six tangent planes form a cube which is circumscribing around the convex set K.

4. Remarks

There are two problems related to our results. The first one is to investigate whether it is possible to inscribe a cube in a given bounded open convex set in R^3. The answer to this question is negative, and a counter-example to this is given by a tetrahedron in R^3 which is extremely flat. In fact, if we take a convex quadrangle $ABCD$ on the (x, y)-plane, such that the two diagonals AC and BD are not perpendicular to each other, and if we shift the vertex A in a direction of z-axis by a small distance, then the tetrahedron $A'BCD$ thus obtained is a required one. It is easy to see that there is no inscribing cube in this tetrahedron.

The second problem concerns the possibility of generalizations to higher dimensional cases. It is not yet known whether or not it is possible to find a circumscribing n-dimensional cube around any given bounded closed convex set in R^n ($n \geqq 4$). We may also ask: Given a real-valued continuous function $f(P)$ defined on an $(n-1)$-sphere S^{n-1}, is it possible to find n points $P_1, \cdots, P_n$ on S^{n-1}, perpendicular to one another, such that $f(P_1) = \cdots = f(P_n)$ ($n \geqq 4$)? These problems are still unsolved.

INSTITUTE FOR ADVANCED STUDY

Annals of Mathematics
Vol. 43, No. 4, October, 1942

AN EXTREMUM PROBLEM IN PRODUCT MEASURE

By Shizuo Kakutani

(Received June 19, 1942)

I. The problem and results

The following problem was proposed by P. R. Halmos: Let Φ be the collection of all real valued measurable functions $\varphi(x, y)$ defined on the unit square $I_x \times I_y$: $0 \leqq x, y \leqq 1$ such that $0 \leqq \varphi(x, y) \leqq 1$ for any $(x, y) \in I_x \times I_y$. Let us put

$$A(\varphi) = \int_0^1 \int_0^1 \varphi(x, y)\, dx\, dy, \tag{1}$$

$$V(\varphi) = \int_0^1 \int_0^1 \int_0^1 \varphi(x, y)\, \varphi(y, z)\, dx\, dy\, dz, \tag{2}$$

and

$$\lambda(\alpha) = \sup_{\substack{A(\varphi)=\alpha \\ \varphi \in \Phi}} V(\varphi), \tag{3}$$

$$\mu(\alpha) = \inf_{\substack{A(\varphi)=\alpha \\ \varphi \in \Phi}} V(\varphi), \tag{4}$$

where α is a real number ($0 \leqq \alpha \leqq 1$). Then what are the exact values of $\lambda(\alpha)$ and $\mu(\alpha)$ as functions of α in the interval $0 \leqq \alpha \leqq 1$?

Consider the special case when $\varphi(x, y)$ is the characteristic function $\varphi_E(x, y)$ of a measurable set $E \subseteq I_x \times I_y$. Then $A(\varphi_E) = A(\varphi)$ is clearly the area (= two dimensional Lebesgue measure) of the set E, while the meaning of $V(\varphi_E) = V(\varphi)$ may be interpreted as follows: Take the unit cube $I_x \times I_y \times I_z$: $0 \leqq x, y, z \leqq 1$, and consider E as a subset of its face $I_x \times I_y \times (0)$. Let E' be the set on the face $(0) \times I_y \times I_z$ which is obtained from E by the mapping $(x, y, 0) \to (0, x, y)$. Then $V(\varphi_E)$ is the volume (= three dimensional Lebesgue measure) of the intersection of two cylindrical sets $E \times I_z$ and $I_x \times E'$, i.e., the set of all points $(x, y, z) \in I_x \times I_y \times I_z$ such that $(x, y) \in E$ and $(y, z) \in E'$.

The purpose of the present note is to prove the following

Theorem.

$$\lambda(\alpha) = 2\alpha - 1 + (1 - \alpha)^{3/2}, \qquad 0 \leqq \alpha \leqq \tfrac{1}{2} \tag{5}$$

$$\lambda(\alpha) = \alpha^{3/2}, \qquad \tfrac{1}{2} \leqq \alpha \leqq 1 \tag{6}$$

$$\begin{gathered} \mu(\alpha) = \frac{n-2}{3n^2}\left\{(3\alpha - 1)\, n + 1 - \frac{((1 - 2\alpha)\, n - 1)^{3/2}}{(n-1)^{1/2}}\right\}, \\ \frac{1}{2}\left(1 - \frac{1}{n-1}\right) \leqq \alpha \leqq \tfrac{1}{2}\left(1 - \frac{1}{n}\right), \qquad n = 2, 3, \cdots \end{gathered} \tag{7}$$

$$\mu(\tfrac{1}{2}) = \tfrac{1}{6}, \tag{8}$$

$$\begin{gathered} \mu(\alpha) = 2\alpha - 1 + \frac{n-2}{3n^2}\left\{(2 - 3\alpha)\, n + 1 - \frac{((2\alpha - 1)\, n - 1)^{3/2}}{(n-1)^{1/2}}\right\}, \\ \frac{1}{2}\left(1 + \frac{1}{n}\right) \leqq \alpha \leqq \frac{1}{2}\left(1 + \frac{1}{n-1}\right), \qquad n = 2, 3, \cdots. \end{gathered} \tag{9}$$

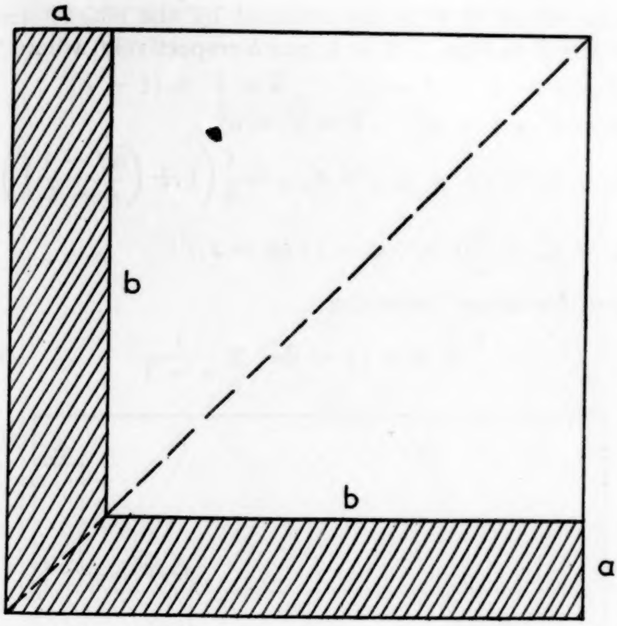

FIG. 1

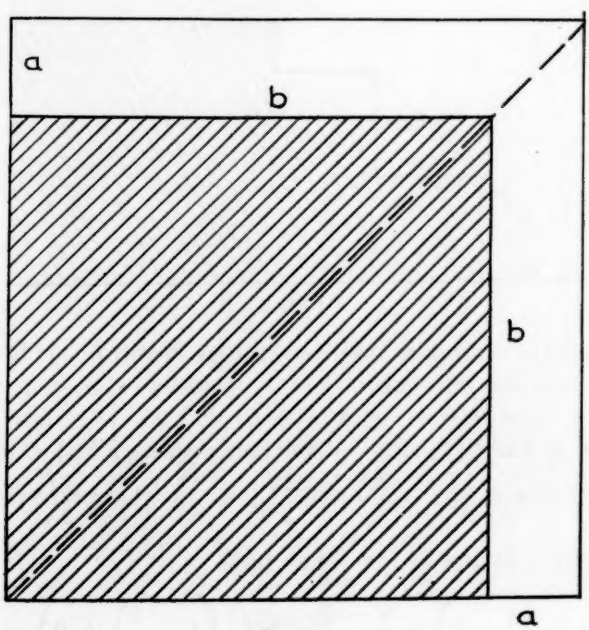

FIG. 2

These extreme values of $V(\varphi)$ are attained by the characteristic functions $\varphi_E(x, y)$ of the sets E of Figs. 1, 2, 3, 4, and 5 respectively, where

$$\bar{a} = \bar{a}' = 1 - (1 - \alpha)^{\frac{1}{2}}, \qquad \bar{b} = \bar{b}' = (1 - \alpha)^{\frac{1}{2}} \qquad \text{(in Fig. 1);}$$

$$\bar{a} = \bar{a}' = 1 - \alpha^{\frac{1}{2}}, \qquad \bar{b} = \bar{b}' = \alpha^{\frac{1}{2}} \qquad \text{(in Fig. 2);}$$

$$\bar{a}_1 = \bar{a}_1' = \cdots = \bar{a}_{n-1} = \bar{a}_{n-1}' = \frac{1}{n}\left(1 + \left(\frac{n\beta - 1}{n - 1}\right)^{\frac{1}{2}}\right)$$

$$\bar{a}_n = \bar{a}_n' = \frac{1}{n}(1 - ((n - 1)(n\beta - 1))^{\frac{1}{2}})$$

where n is a positive integer satisfying

$$\frac{1}{n} \leqq \beta \equiv |1 - 2\alpha| \leqq \frac{1}{n - 1} \qquad \text{(in Figs. 3, 5).}$$

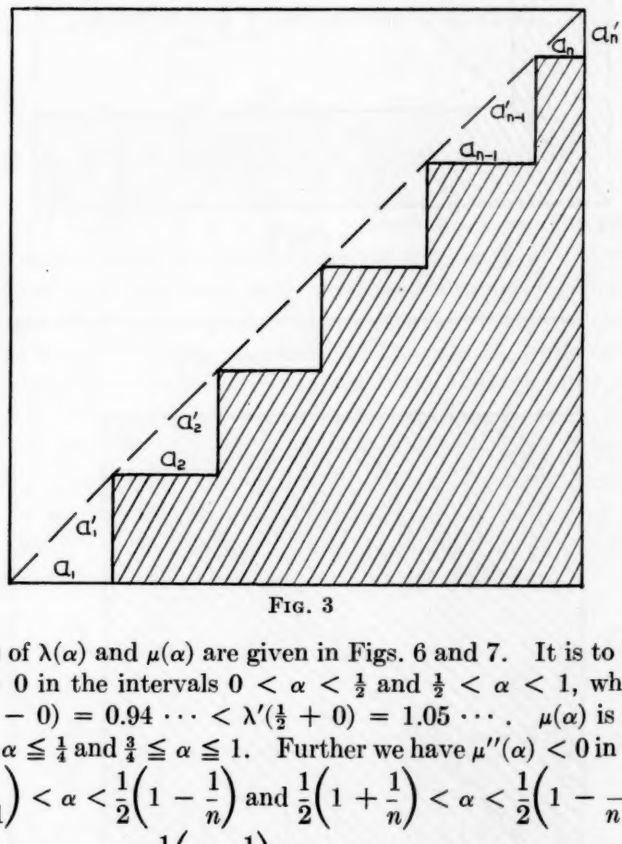

FIG. 3

The graphs of $\lambda(\alpha)$ and $\mu(\alpha)$ are given in Figs. 6 and 7. It is to be remarked that $\lambda''(\alpha) > 0$ in the intervals $0 < \alpha < \frac{1}{2}$ and $\frac{1}{2} < \alpha < 1$, while at $\alpha = \frac{1}{2}$ we have $\lambda'(\frac{1}{2} - 0) = 0.94 \cdots < \lambda'(\frac{1}{2} + 0) = 1.05 \cdots$. $\mu(\alpha)$ is linear in the intervals $0 \leqq \alpha \leqq \frac{1}{4}$ and $\frac{3}{4} \leqq \alpha \leqq 1$. Further we have $\mu''(\alpha) < 0$ in the intervals $\frac{1}{2}\left(1 - \frac{1}{n - 1}\right) < \alpha < \frac{1}{2}\left(1 - \frac{1}{n}\right)$ and $\frac{1}{2}\left(1 + \frac{1}{n}\right) < \alpha < \frac{1}{2}\left(1 - \frac{1}{n - 1}\right)$, $n = 3, 4, \cdots$. Finally, at $\alpha = \frac{1}{2}\left(1 \pm \frac{1}{n}\right)$ we have

$$\mu'\left(\frac{1}{2}\left(1 - \frac{1}{n}\right) - 0\right) = \frac{n - 2}{n} < \mu'\left(\frac{1}{2}\left(1 - \frac{1}{n}\right) + 0\right) = \frac{n - 1}{n},$$

$$\mu'\left(\frac{1}{2}\left(1 + \frac{1}{n}\right) - 0\right) = \frac{n + 1}{n} < \mu'\left(\frac{1}{2}\left(1 + \frac{1}{n}\right) + 0\right) = \frac{n + 2}{n}.$$

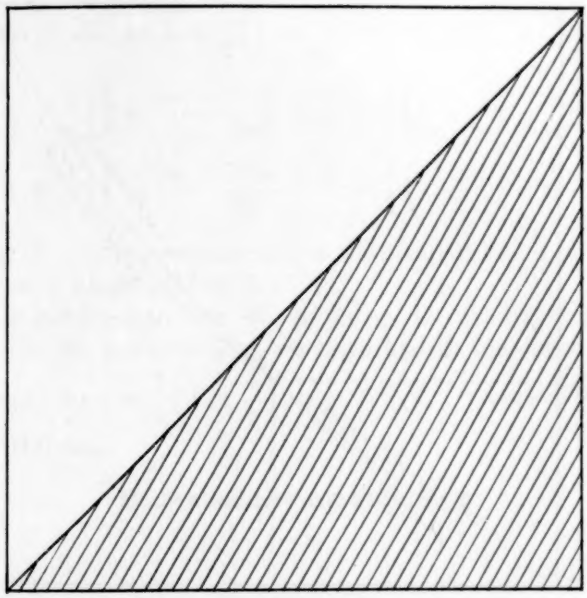

FIG. 4

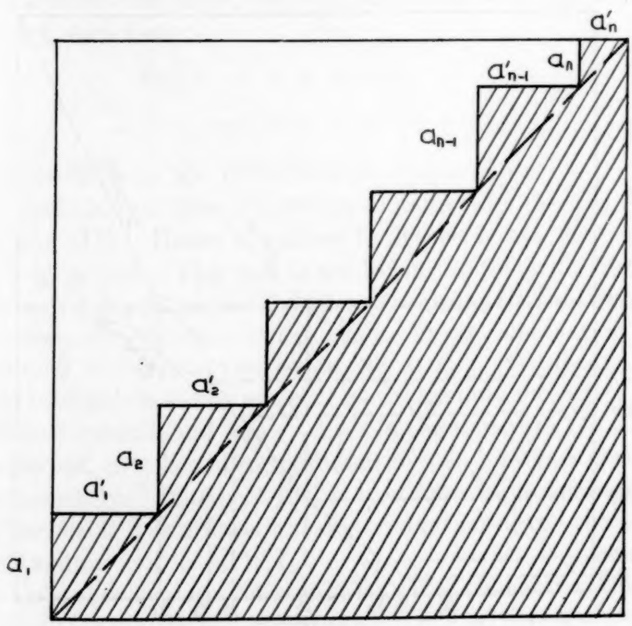

FIG. 5

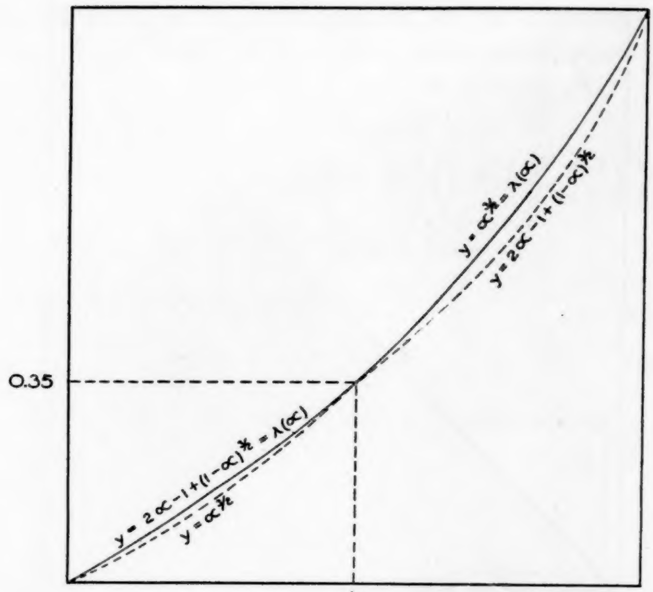
y = α^{3/2} = λ(α)
y = 2α − 1 + (1 − α)^{3/2}
0.35
y = 2α − 1 + (1 − α)^{3/2} = λ(α)
y = α^{3/2}
1/2

FIG. 6

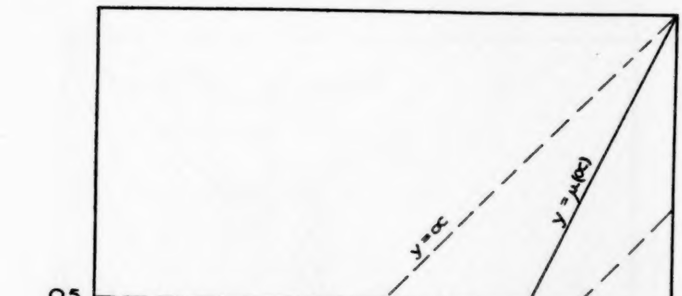
y = α
y = μ(α)
y = α − 1/3
0.5
0.37
0.31
1/6 = 0.16
0.06
0.03
1/4
1/3
3/8
1/2
5/8
2/3
3/4

FIG. 7

The exact values of $\mu(\alpha)$ at $\alpha = \frac{1}{2}\left(1 \pm \frac{1}{n}\right)$, $n = 1, 2, 3, \cdots$, are given by

$$\mu\left(\frac{1}{2}\left(1 - \frac{1}{n}\right)\right) = \frac{(n-1)(n-2)}{6n^2} = \frac{1}{2}\left(1 - \frac{1}{n}\right) - \frac{1}{3} + \frac{1}{3n^2}.$$

$$\mu\left(\frac{1}{2}\left(1 + \frac{1}{n}\right)\right) = \frac{(n+1)(n+2)}{6n^2} = \frac{1}{2}\left(1 + \frac{1}{n}\right) - \frac{1}{3} + \frac{1}{3n^2}.$$

Thus the graph of $\mu(\alpha)$ lies entirely above the straight line $y = \alpha - \frac{1}{3}$, except at the point $\alpha = \frac{1}{2}$ where $\mu(\frac{1}{2}) = \frac{1}{6}$.

The author is indebted to Drs. W. Ambrose, D. Blackwell, R. H. Fox and P. R. Halmos for the conversations we have had in the course of this work. The values of $\mu(\alpha)$ for $\alpha = \frac{1}{2}\left(1 \pm \frac{1}{n}\right)$, $n = 1, 2, 3, \cdots$ were obtained by R. H. Fox and P. R. Halmos.

II. Preliminary considerations

Lemma 1.

$$\lambda(1 - \alpha) = 1 - 2\alpha + \lambda(\alpha), \tag{10}$$

$$\mu(1 - \alpha) = 1 - 2\alpha + \mu(\alpha). \tag{11}$$

Proof. These are the direct consequences of the fact that $\varphi(x, y) \in \Phi$ implies $1 - \varphi(x, y) \in \Phi$, and that

$$A(1 - \varphi) = 1 - A(\varphi), \tag{12}$$

$$V(1 - \varphi) = 1 - 2A(\varphi) + V(\varphi), \tag{13}$$

which follow easily from the definitions of $A(\varphi)$ and $V(\varphi)$.

It is clear that the functions $\lambda(\alpha)$ and $\mu(\alpha)$ defined by (5), (6), (7), (8) and (9) satisfy (10) and (11). Hence it suffices to discuss either the case $0 \leqq \alpha \leqq \frac{1}{2}$ or the case $\frac{1}{2} \leqq \alpha \leqq 1$. This fact is needed in the following discussions.

Definition 1. A real valued function $f(x)$ defined on the interval $I_x : 0 \leqq x \leqq 1$ is an *elementary function* if it is a finite linear combination of the characteristic functions of the intervals contained in I_x. (We do not care whether each of these intervals is closed or open, as we are only interested in elementary functions defined modulo null sets.) A set in the square $I_x \times I_y : 0 \leqq x, y \leqq 1$ is a *rectangular set*, if it is a direct product of two intervals each contained in I_x and I_y respectively. Finally, a real valued function defined on $I_x \times I_y$ is an *elementary function* if it is a finite linear combination of the characteristic functions of rectangular sets in $I_x \times I_y$.

Let Φ^0 be the subcollection of Φ consisting of all elementary functions $\varphi(x, y)$ in Φ. It is clear that in the definitions (3), (4) of $\lambda(\alpha)$ and $\mu(\alpha)$, we may replace the condition $\varphi \in \Phi$ by $\varphi \in \Phi^0$ and yet we obtain the same sup or inf. Hence, in order to prove our theorem, it suffices to show that $\mu(\alpha) \leqq V(\varphi) \leqq \lambda(\alpha)$

for any $\varphi(x, y) \in \Phi^0$ with $A(\varphi) = \alpha$, where $\lambda(\alpha)$ and $\mu(\alpha)$ are defined by (5), (6), (7), (8) and (9).

Let now $\varphi(x, y) \in \Phi^0$. We shall put

$$f_\varphi(x) = \int_0^1 \varphi(x, y)\, dy, \tag{14}$$

$$g_\varphi(y) = \int_0^1 \varphi(x, y)\, dx. \tag{15}$$

It is clear that $f_\varphi(x)$ and $g_\varphi(y)$ are elementary functions and we have

$$A(\varphi) = \int_0^1 f_\varphi(t)\, dt = \int_0^1 g_\varphi(t)\, dt, \tag{16}$$

$$V(\varphi) = \int_0^1 f_\varphi(t) g_\varphi(t)\, dt. \tag{17}$$

LEMMA 2. *For any $\varphi(x, y) \in \Phi^0$, there exists a $\varphi'(x, y) \in \Phi^0$ such that $A(\varphi') = A(\varphi)$, $V(\varphi') = V(\varphi)$, and such that $f_{\varphi'}(x)$ is monotone non-increasing or monotone non-decreasing in x.*

PROOF. Since $f_\varphi(x)$ is an elementary function, there exists a measure preserving transformation $x' = h(x)$ of the interval I_x onto itself, such that $f_\varphi(h(x))$ is monotone non-increasing or monotone non-decreasing. In fact, we can choose $h(x)$ as a permutation of subintervals of I_x. It is then clear that the function $\varphi'(x, y) = \varphi(h(x), h(y))$ satisfies all the conditions required in Lemma 2.

DEFINITION 2. A set $E \subset I_x \times I_y$ is an *elementary set* if it is a union of a finite number of rectangular sets. A set $E \subset I_x \times I_y$ is a *corner set* if $(x, y) \in E$, $0 \leqq x' \leqq x$, $0 \leqq y' \leqq y$ imply $(x', y') \in E$. Further, a set $E \subset I_x \times I_y$ is a *corner* set* if $(x, y) \in E$, $x \leqq x' \leqq 1$, $0 \leqq y' \leqq y$ imply $(x', y') \in E$.

LEMMA 3. *Let $\varphi(x, y) \in \Phi^0$ be an elementary function such that $f_\varphi(x)$ is monotone non-increasing in x. Then there exists an elementary corner set $E \subset I_x \times I_y$ such that $A(\varphi_E) = A(\varphi)$, $V(\varphi_E) \geqq V(\varphi)$.*

PROOF. Let E be the set of all points $(x, y) \in I_x \times I_y$ such that $0 \leqq y \leqq f(x)$. It is clear that E is an elementary corner set satisfying

$$f_{\varphi_E}(t) = f_\varphi(t), \quad \text{for} \quad 0 \leqq t \leqq 1. \tag{18}$$

From this follows easily that $A(\varphi_E) = A(\varphi)$. Moreover, it is easy to see that

$$G_{\varphi_E}(t) \geqq G_\varphi(t), \quad \text{for} \quad 0 \leqq t \leqq 1, \tag{19}$$

$$G_{\varphi_E}(0) = G_\varphi(0)(= 0), \qquad G_{\varphi_E}(1) = G_\varphi(1)(= A(\varphi_E) = A(\varphi)), \tag{20}$$

where $G_{\varphi_E}(t)$ and $G_\varphi(t)$ are defined by

$$G_{\varphi_E}(t) = \int_0^t g_{\varphi_E}(s)\, ds, \qquad G_\varphi(t) = \int_0^t g_\varphi(s)\, ds, \qquad \text{for } 0 \leqq t \leqq 1 \tag{21}$$

respectively.

Consequently,

$$
\begin{aligned}
V(\varphi_E) &= \int_0^1 f_{\varphi_E}(t) g_{\varphi_E}(t)\, dt \\
&= [f_{\varphi_E}(t) G_{\varphi_E}(t)]_0^1 - \int_0^1 G_{\varphi_E}(t)\, df_{\varphi_E}(t) \\
&\geqq [f_\varphi(t) G_\varphi(t)]_0^1 - \int_0^1 G_\varphi(t)\, df_\varphi(t) \\
&= \int_0^1 f_\varphi(t) g_\varphi(t)\, dt = V(\varphi),
\end{aligned} \tag{22}
$$

which proves Lemma 3.

In the same way we can prove

LEMMA 3′. *Let $\varphi(x, y) \in \Phi^0$ be an elementary function such that $f_\varphi(x)$ is monotone non-decreasing in x. Then there exists an elementary corner* set $E \subset I_x \times I_y$ such that $A(\varphi_E) = A(\varphi)$, $V(\varphi_E) \leqq V(\varphi)$.*

We omit the proof.

DEFINITION 3. Let E be an elementary corner set in $I_x \times I_y$, and consider the graphs $y = f_{\varphi_E}(x)$ and $x = g_{\varphi_E}(y)$. These two graphs together will compose a polygonal line Γ_E, consisting only of horizontal and vertical segments, which connects the points (0, 1) and (1, 0). This polygonal line is called the *characteristic graph* of E. Similarly, we can define the characteristic graph of an elementary corner* set $E \subset I_x \times I_y$. This is a polygonal line connecting two points (0, 0) and (1, 1).

We shall divide our further arguments into two parts, namely, the discussion of $\lambda(\alpha)$ and that of $\mu(\alpha)$.

III. Discussion of $\lambda(\alpha)$

DEFINITION 4. A set $E \subset I_x \times I_y$ is *symmetric* if $(x, y) \in E$ implies $(y, x) \in E$.

LEMMA 4. *For any elementary corner set $E \subset I_x \times I_y$, there exists a symmetric elementary corner set $E' \subset I_x \times I_y$ such that $A(\varphi_{E'}) = A(\varphi_E)$, $V(\varphi_{E'}) \geqq V(\varphi_E)$.*

PROOF. Let Γ_E be the characteristic graph of E. Γ_E has a unique intersection with the diagonal $x = y$ of the unit square $I_x \times I_y$. Let (ξ, ξ) be this point of intersection. Then the required set E' is defined as the set of all points $(x, y) \in I_x \times I_y$ satisfying one of the following three conditions:

$$0 \leqq x \leqq \xi, \quad 0 \leqq y \leqq \xi, \tag{23}$$

$$0 \leqq x \leqq \tfrac{1}{2}\{f_{\varphi_E}(y) + g_{\varphi_E}(y)\}, \quad \xi < y \leqq 1, \tag{24}$$

$$\xi < x \leqq 1, \quad 0 \leqq y \leqq \tfrac{1}{2}\{f_{\varphi_E}(x) + g_{\varphi_E}(y)\}. \tag{25}$$

It is clear that E' is a symmetric elementary corner set, and that $A(\varphi_{E'}) = A(\varphi_E)$. In order to prove that $V(\varphi_{E'}) \geqq V(\varphi_E)$, we put

(26) $p(t) = f_{\varphi_E}(t), \quad q(t) = g_{\varphi_E}(t), \quad r(t) = f_{\varphi_{E'}}(t) = g_{\varphi_{E'}}(t), \quad \text{for } 0 \leqq t \leqq 1,$

(27)
$$P(t) = \int_0^t p(s)\, ds, \qquad Q(t) = \int_0^t q(s)\, ds, \qquad R(t) = \int_0^t r(s)\, ds,$$
for $0 \leqq t \leqq 1$.

It is then easy to see that

(28) $$2R(t) \geqq P(t) + Q(t), \quad \text{for} \quad 0 \leqq t \leqq \xi,$$

(29) $$2r(t) = p(t) + q(t), \quad \text{for} \quad \xi < t \leqq 1.$$

Consequently,

(30)
$$\begin{aligned}
\int_0^\xi \{r(t)^2 - p(t)q(t)\}\, dt &\geqq \frac{1}{4}\int_0^\xi \{4r(t)^2 - (p(t) + q(t))^2\}\, dt \\
&= \frac{1}{4}\int_0^\xi \{2r(t) - p(t) - q(t)\}\{2r(t) + p(t) + q(t)\}\, dt \\
&= \frac{1}{4}[\{2R(t) - P(t) - Q(t)\}\{2r(t) + p(t) + q(t)\}]_0^\xi \\
&\quad - \frac{1}{4}\int_0^\xi \{2R(t) - P(t) - Q(t)\}\, d\{2r(t) + p(t) + q(t)\}\, dt \geqq 0
\end{aligned}$$

and

(31)
$$\begin{aligned}
\int_\xi^1 \{r(t)^2 - p(t)q(t)\}\, dt &= \frac{1}{4}\int_\xi^1 \{(p(t) + q(t))^2 - p(t)q(t)\}\, dt \\
&= \frac{1}{4}\int_0^1 (p(t) - q(t))^2\, dt \geqq 0,
\end{aligned}$$

which together imply

(32) $$V(\varphi_{E'}) - V(\varphi_E) = \int_0^1 \{r(t)^2 - p(t)q(t)\}\, dt \geqq 0.$$

The proof of Lemma 4 is completed.

Thus in order to discuss $\lambda(\alpha)$, it suffices to consider symmetric elementary corner sets $E \subset I_x \times I_y$.

Now let E be a symmetric elementary corner set in $I_x \times I_y$, and let Γ_E be its characteristic graph. Let us assume that Γ_E contains at least three segments above the diagonal $x = y$.[1] Let a, b, c be three consecutive segments of Γ_E lying above the diagonal $x = y$. We assume that a and c are horizontal, while b is vertical to the x-axis. (See Fig. 8.) We shall replace the part of Γ_E consisting of a, b, c by another system of segments a', b', c' as indicated in Fig. 8. Let us denote the new symmetric elementary corner set thus obtained by E'. (Of course, we make the same change below the diagonal, as indicated in Fig. 6,

[1] When we say that a segment (parallel to the x-axis or to the y-axis) lies above or below the diagonal $x = y$, one of the end points of the segment may lie on the diagonal $x = y$, while on the other hand, when we say that a segment lies *entirely* above or below the diagonal neither of the end points of the segment can lie on the diagonal $x = y$.

so as to make E' symmetric.) The y-coordinate of b' is so chosen that we have $A(\varphi_{E'}) = A(\varphi_E)$, and this condition is thus fulfilled if we have $\bar{a}/\bar{b}' = \bar{c}'/\bar{b} = \theta$, where θ is a real number satisfying $0 < \theta < 1$. We shall compare $V(\varphi_{E'})$ with $V(\varphi_E)$. A simple computation shows

$$V(\varphi_{E'}) - V(\varphi_E) = \theta(1 - \theta)(\bar{a} + \bar{c})\bar{b}(\bar{a} + \bar{c} - \bar{b}). \tag{33}$$

Hence, if $\bar{a} + \bar{c} > \bar{b}$, then by replacing a, b, c by a', b', c' we obtain a new symmetric elementary corner set E' such that $A(\varphi_{E'}) = A(\varphi_E)$, $V(\varphi_{E'}) \geqq V(\varphi_E)$. If we interchange a, b, c with a', b', c', then we immediately see that the same thing is true even if a and b are vertical, while b is horizontal to the x-axis.

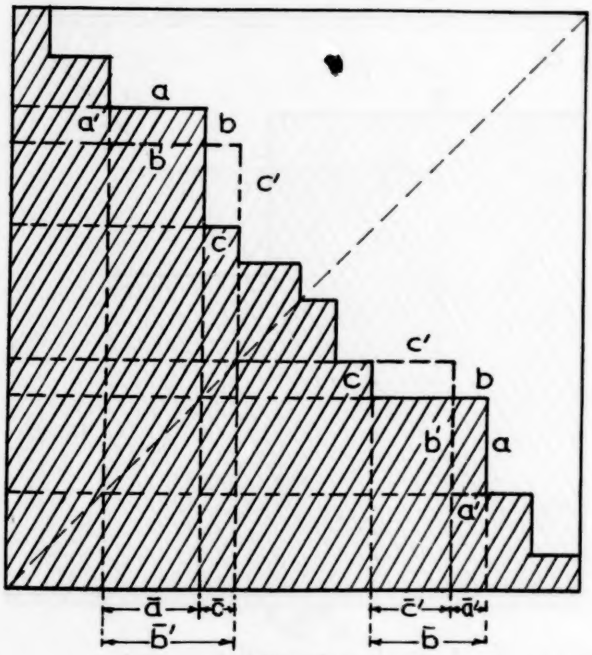

FIG. 8

Assume now that Γ_E contains at least four segments above the diagonal $x = y$. Let a, b, c, d be any four consecutive segments. We denote their respective lengths by $\bar{a}, \bar{b}, \bar{c}, \bar{d}$. It is then easy to see that at least one of the inequalities $\bar{a} + \bar{c} > \bar{b}$, $\bar{b} + \bar{d} > \bar{c}$ must hold. Hence, by replacing a, b, c or b, c, d by a suitable system a', b', c' or b', c', d', we can always obtain a new symmetric elementary corner set E' for which $A(\varphi_{E'}) = A(\varphi_E)$, $V(\varphi_{E'}) \geqq V(\varphi_E)$. Further, it is to be noticed that the number of segments lying above the diagonal $x = y$ in the characteristic graph of the new set E' is smaller than that of E exactly by two.

Thus, by iterating the same process, we shall finally reach a symmetric elementary corner set E^* whose characteristic graph Γ_{E^*} consists of at most three segments lying above the diagonal $x = y$, and such that $A(\varphi_{E^*}) = A(\varphi_E)$,

$V(\varphi_{E^*}) \geqq V(\varphi_E)$. Consequently, in order to discuss $\lambda(\alpha)$ it suffices to consider the symmetric elementary sets E of the forms given in Figs. 1, 2, 9 and 10.

We shall discuss these cases separately.

(i) CASE OF FIG. 1. The condition $A(\varphi_E) = \alpha$ implies $\bar{a} = 1 - (1 - \alpha)^{\frac{1}{2}}$, $\bar{b} = (1 - \alpha)^{\frac{1}{2}}$. Consequently

$$V(\varphi_E) = 2\alpha - 1 + (1 - \alpha)^{\frac{3}{2}}. \tag{34}$$

(ii) CASE OF FIG. 2. The condition $A(\varphi_E) = \alpha$ implies $\bar{a} = 1 - \alpha^{\frac{1}{2}}$, $\bar{b} = \alpha^{\frac{1}{2}}$. Consequently,

$$V(\varphi_E) = \alpha^{\frac{3}{2}}. \tag{35}$$

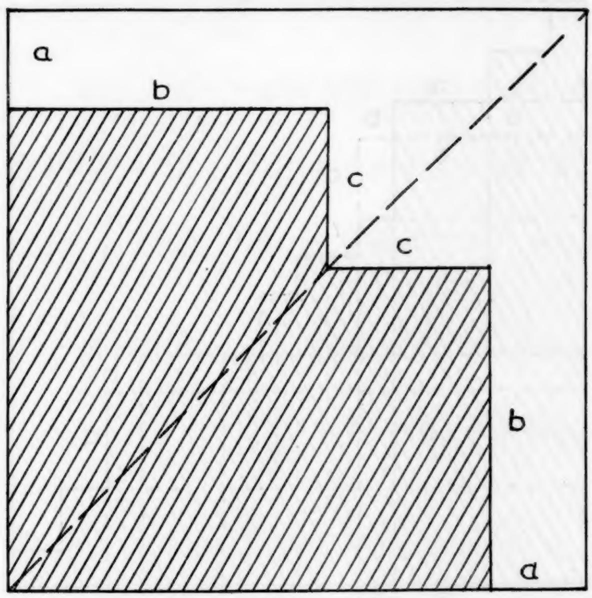

FIG. 9

(iii) CASE OF FIG. 9. A simple computation shows:

$$A(\varphi_E) = 2\bar{b}\bar{c} + \bar{b}^2 = \alpha, \tag{36}$$

$$V(\varphi_E) = \bar{b}(\bar{b} + \bar{c})^2 + \bar{b}^2 c. \tag{37}$$

Hence $c = (\alpha - \bar{b}^2)/2\bar{b}$. Putting this value in (37), we have

$$\begin{aligned} V(\varphi_E) &= \frac{1}{4\bar{b}}(\alpha^2 + 4\alpha\bar{b}^2 - \bar{b}^4) \\ &= \frac{1}{4\bar{b}}\{2\alpha^2 + 2\alpha\bar{b}^2 - (\alpha - \bar{b}^2)^2\} \\ &\leqq \frac{1}{4\bar{b}}(2\alpha^2 + 2\alpha b^2) = \frac{\alpha}{2}\left(\frac{\alpha}{\bar{b}} + \bar{b}\right) \\ &\leqq \frac{\alpha}{2}\cdot 2\left(\frac{\alpha}{\bar{b}}\cdot\bar{b}\right)^{\frac{1}{2}} = \alpha^{\frac{3}{2}}. \end{aligned} \tag{38}$$

(iv) CASE OF FIG. 10. A simple computation shows:

$$A(\varphi_E) = 1 - (2\bar{b}\bar{c} + \bar{b}^2) = \alpha, \tag{39}$$

$$\begin{aligned} V(\varphi_E) &= \bar{a} + \bar{c}(\bar{a} + \bar{c})^2 + \bar{a}^2\bar{b} \\ &= 2\alpha - 1 + \{\bar{b}(\bar{b} + \bar{c})^2 + \bar{b}^2\bar{c}\}. \end{aligned} \tag{40}$$

Consequently, by the result obtained above in the case of Fig. 7,

$$V(\varphi_E) \leqq 2\alpha - 1 + (1 - \alpha)^{\frac{3}{2}}. \tag{41}$$

Summing up, we have thus proved that

$$V(\varphi) \leqq \max\,(\alpha^{\frac{3}{2}}, 2\alpha - 1 + (1 - \alpha)^{\frac{3}{2}}) \tag{42}$$

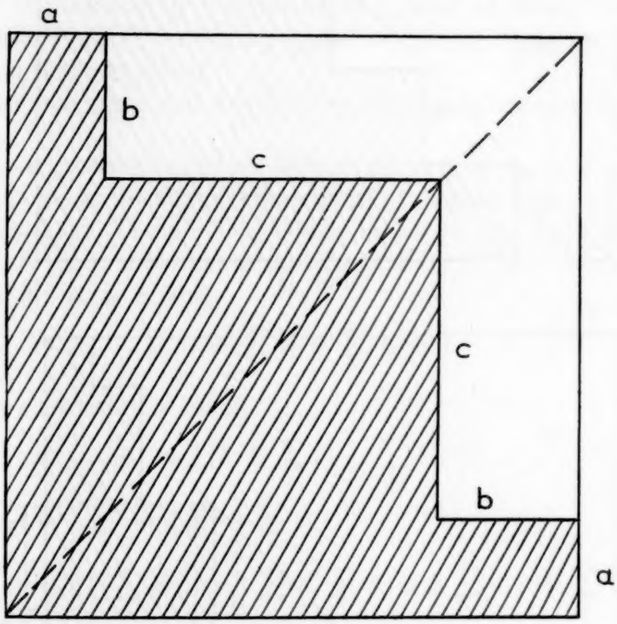

FIG. 10

for any $\varphi(x, y) \in \Phi$ with $A(\varphi) = \alpha$, $0 \leqq \alpha \leqq 1$, the equality holding for the symmetric elementary corner sets E of the forms given in Figs. 1 and 2. This completes the proof of our theorem for $\lambda(\alpha)$.

IV. Discussion of $\mu(\alpha)$

DEFINITION 5. Let E be an elementary corner* set in $I_x \times I_y$, and let Γ_E be its characteristic graph. E is a *special* elementary corner* set if there is no segment in Γ_E which lies *entirely*[2] above or below the diagonal $x = y$. For example, Fig. 11 shows a special elementary corner* set, while this is not the case in Fig. 12.

LEMMA 5. *For any elementary corner* set $E \subset I_x \times I_y$, there exists a special elementary corner* set $E' \subset I_x \times I_y$, such that $A(\varphi_{E'}) = A(\varphi_E)$, $V(\varphi_{E'}) = V(\varphi_E)$.*

[2] See footnote (1) on page 750.

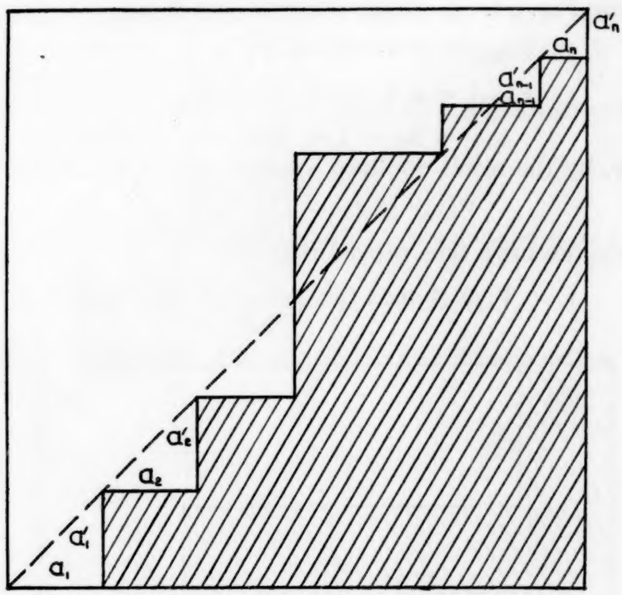

FIG. 11

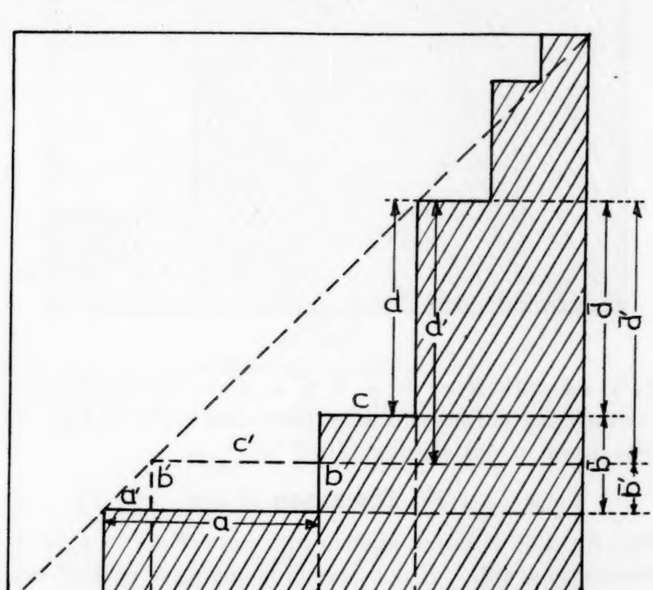

FIG. 12

PROOF. Let us assume that the given elementary corner* set E is of the form as given in Fig. 12. We shall replace the part of Γ_E consisting of a, b, c, d

by a', b', c', d' as indicated in Fig. 12. (Clearly we have $\bar{a} + \bar{c} = \bar{b} + \bar{d} = \bar{a}' + \bar{c}' = \bar{b}' + \bar{d}'$. Let us denote the new set thus obtained by E'. The condition $A(\varphi_{E'}) = A(\varphi_E)$ is fulfilled by taking $\bar{b}\bar{c} = \bar{b}'\bar{c}'$. Then a simple computation shows that $V(\varphi_{E'}) = V(\varphi_E)$.

Thus our lemma is proved if E is an elementary corner* set E of the form of Fig. 12. The analogous argument applies to the case when similar situation happens above the diagonal $x = y$. Finally, if there are more than two (and hence $\geqq 4$) consecutive segments or more than one pair of segments which lie entirely above or below the diagonal $x = y$, then we can iterate the same kind of operations, reducing the number of segments lying entirely above or below the diagonal $x = y$ exactly by two in each step, until we finally reach a special elementary corner* set E^* satisfying $A(\varphi_{E^*}) = A(\varphi_E)$, $V(\varphi_{E^*}) = V(\varphi_E)$. The proof of Lemma 5 is completed.

Thus, in order to discuss $\mu(\alpha)$ it suffices to consider special elementary corner* sets only.

Let now E be a special elementary corner* set in $I_x \times I_y$, as in Fig. 11. We denote the segments of its characteristic graph Γ_E successively by $a_1, a_1', \cdots, a_n, a_n'$ (see Fig. 11). Those a_i, a_i' which lie above the diagonal $x = y$ are denoted by b_j, b_j', and those below the diagonal by c_k, c_k'. We have clearly, $\bar{a}_i = \bar{a}_i'$, $\bar{b}_j = \bar{b}_j'$, $\bar{c}_k = \bar{c}_k'$ and

$$\sum_i \bar{a}_i = \sum_j \bar{b}_j + \sum_k \bar{c}_k = 1. \tag{43}$$

Then a simple computation shows

$$A(\varphi_E) = \tfrac{1}{2}\{1 + \sum_j \bar{b}_j^2 - \sum_k \bar{c}_k^2\} = \alpha, \tag{44}$$

$$\begin{aligned} V(\varphi_E) &= \sum_{i<j<k} \bar{a}_i \bar{a}_j \bar{a}_k + \sum_j \bar{b}_j \\ &= \tfrac{1}{6}\{(\sum_i \bar{a}_i)^3 - 3 \sum_i \bar{a}_i^2 \sum_i \bar{a}_i + 2 \sum_i \bar{a}_i^3\} + \sum_j \bar{b}_j^2 \\ &= \tfrac{1}{6}\{1 - 3 \sum_i \bar{a}_i^2 + 2 \sum_i \bar{a}_i^3\} + \sum_j \bar{b}_j^2 \\ &= \tfrac{1}{6}\{1 + 3(\sum_j \bar{b}_j^2 - \sum_k \bar{c}_k^2) + 2 \sum_i \bar{a}_i^3\} \\ &= \tfrac{1}{6}\{1 + 3(2\alpha - 1) + 2 \sum_i \bar{a}_i^3\} \\ &= \alpha - \tfrac{1}{3} + \tfrac{1}{3}\{\sum_j \bar{b}_j^3 + \sum_k \bar{c}_k^3\}. \end{aligned} \tag{45}$$

Thus our problem is transformed into the following one: *under the conditions* (43) *and*

$$\sum_j \bar{b}_j^2 - \sum_k \bar{c}_k^2 = 2\alpha - 1 \tag{46}$$

to make

$$\sum_j \bar{b}_j^3 + \sum_k \bar{c}_k^3 = \omega \equiv 3(V(\varphi_E) - \alpha) + 1 \tag{47}$$

as small as possible, where $\bar{b}_j \geqq 0$ *and* $\bar{c}_k \geqq 0$ *and there is no assumption on the number of* $\bar{b}_j$ *and* $\bar{c}_k$.

Let us consider the interval $0 \leqq \alpha \leqq \frac{1}{2}$. Then it is clear that we have only to consider the case when all $\bar{b}_j = 0$, and our problem is further reduced to the following one: *Under the conditions*:

$$\sum_k \bar{c}_k = 1, \tag{48}$$

$$\sum_k \bar{c}_k^2 = \beta \equiv 1 - 2\alpha > 0, \tag{49}$$

to make

$$\sum_k \bar{c}_k^3 = \omega = 3(V(\varphi_E) - \alpha) + 1 \tag{50}$$

as small as possible, where $\bar{c}_k \geqq 0$ *and we have no assumption on the number n of* c_k.

By Schwarz's inequality, we have $n\beta > 1$, and it is easy to see that for fixed n with $n\beta \geqq 1$, the minimum value ω_n of ω is attained by

$$c_1 = \cdots = c_{n-1} = \frac{1}{n}\left(1 + \left(\frac{n\beta - 1}{n - 1}\right)^{\frac{1}{2}}\right), \tag{51}$$

$$c_n = \frac{1}{n}\left(1 - ((n\beta - 1)(n - 1))^{\frac{1}{2}}\right) \tag{52}$$

and

$$\omega_n = \frac{1}{n^2}\left\{(3n\beta - 2) - \frac{(n - 2)(n\beta - 1)^{\frac{3}{2}}}{(n - 1)^{\frac{1}{2}}}\right\}. \tag{52}$$

Hence, by (50) and (51), we finally have

$$\begin{aligned} V(\varphi_E) &= \alpha - \frac{1}{3} + \frac{1}{3n^2}\left\{(3n(1 - 2\alpha) - 2) - \frac{(n - 2)(n(1 - 2\alpha) - 1)^{\frac{3}{2}}}{(n - 1)^{\frac{1}{2}}}\right\} \\ &= \frac{n - 2}{3n^2}\left\{(3\alpha - 1)n + 1 - \frac{(n(1 - 2\alpha) - 1)^{\frac{3}{2}}}{(n - 1)^{\frac{1}{2}}}\right\} \end{aligned} \tag{53}$$

where φ_E is the characteristic function of a special elementary corner* set E of the form given in Fig. 3. It is easy to see that the expression (54) is a monotone increasing function of n for each given α for $n \geqq (1 - 2\alpha)^{-1}$. Hence the smallest possible integer with $n\beta \equiv n(1 - 2\alpha) \geqq 1$ gives the required value of $\mu(\alpha)$, or in other words, the equality (7) is true for $\alpha \geqq \frac{1}{2}\left(1 - \frac{1}{n - 1}\right), < \frac{1}{2}\left(1 - \frac{1}{n}\right)$. Thus we have proved the formula (7) for $n = 2, 3, \cdots$, i.e. for all α satisfying $0 \leqq \alpha < \frac{1}{2}$. The formula (9) for $n = 2, 3, \cdots$, or for $\frac{1}{2} < \alpha \leqq 1$ then follows from this and from Lemma 1. Finally, the formula (8) for $\alpha = \frac{1}{2}$ follows from (7), (9) and from the fact that $\mu(\alpha)$ is a monotone non-decreasing function of α. This completes the discussion of $\mu(\alpha)$.

INSTITUTE FOR ADVANCED STUDY

Annals of Mathematics
Vol. 43, No. 4, October, 1942

GROUP EXTENSIONS AND HOMOLOGY*

By Samuel Eilenberg and Saunders MacLane

(Received May 21, 1942)

CONTENTS

* Presented to the American Mathematical Society, September 4 and December 31, 1941. Part of the results was published in a preliminary report [5] and also in an appendix to Lefschetz [7]. The numbers in brackets refer to the bibliography at the end of the paper.

INTRODUCTION

In 1937 the following problem was formulated by Borsuk and Eilenberg: Given a solenoid[1] Σ in the three sphere S^3, how many homotopy classes of continuous mappings $f(S^3 - \Sigma) \subset S^2$ are there? In 1939 Eilenberg proved ([4], p. 251) that the homotopy classes in question are in a 1-1-correspondence with the elements of the one-dimensional homology group $H^1(K, I) = Z^1(K, I)/B^1(K, I)$, where K is any representation of $S^3 - \Sigma$ as a complex, $Z^1(K, I)$ is the group of infinite 1-cycles in K with the additive group I of integers as coefficients and $B^1(K, I)$ is the subgroup of bounding cycles. This homology group is generally much "larger" than the conventional homology group $H_f^1(K, I) = Z^1/\bar{B}^1$ where $\bar{B}^1(K, I)$ is the group of cycles that bound on every finite portion of K; with an appropriate topology in the group Z^1, $\bar{B}^1$ turns out to be exactly the closure of B^1.

At this point the investigation was taken up by Steenrod [10]. By using "regular cycles" he computed the groups $H^1(S^3 - \Sigma)$ for the various solenoids Σ. The groups are uncountable and of a rather complicated nature.[2]

This paper originated from an accidental observation that the groups obtained by Steenrod were identical with some groups that occur in the purely algebraic theory of *extensions of groups*. An abelian group E is called an ex-

[1] For the definition see Appendix B below.

[2] A popular exposition of Steenrod's results can be found in his article in Lectures in Topology, Ann Arbor, University of Michigan Press, 1941, pp. 43–55.

tension of the group G by the group H if $G \subset E$ and $H = E/G$. With a proper definition of equivalence and addition, the extensions of G by H themselves form an abelian group Ext $\{G, H\}$. It turns out that $H^1(S^3 - \Sigma, I)$ is isomorphic with Ext $\{I, \Sigma^*\}$ where Σ^* is a properly chosen subgroup of the group of rational numbers.[3]

The thesis of this paper is that the theory of group extensions forms a natural and powerful tool in the study of homologies in infinite complexes and topological spaces. Even in the simple and familiar case of finite complexes the results obtained are finer than the existing ones.

Our fundamental theorem concerns the homology groups of a star finite complex K. Let $H^q(G)$ denote the homology group of infinite cycles with coefficients in an arbitrary topological group G. We obtain an explicit expression for $H^q(G)$ in terms of G and the cohomology groups $\mathcal{H}_q$ of *finite* cocycles with integral coefficients. ($\mathcal{H}_q$ is the factor group $\mathcal{Z}_q/\mathcal{B}_q$ of cocycles modulo coboundaries). This expression is

$$H^q(G) = \text{Hom}\,\{\mathcal{H}_q, G\} \times \text{Hom}\,\{\mathcal{B}_{q+1}, G\}/\text{Hom}\,\{\mathcal{Z}_{q+1} \mid \mathcal{B}_{q+1}, G\}.$$

Here Hom $\{H, G\}$ stands for the (topological) group of all homomorphisms of H into G, while Hom $\{\mathcal{Z}_{q+1} \mid \mathcal{B}_{q+1}, G\}$ denotes the group of those homomorphisms of $\mathcal{B}_{q+1}$ into G which can be extended to homomorphisms of $\mathcal{Z}_{q+1}$ into G. The factor group on the right in this expression appears to depend on the groups $\mathcal{B}_{q+1}$ and $\mathcal{Z}_{q+1}$, but actually depends only on the cohomology group $\mathcal{H}_{q+1} = \mathcal{Z}_{q+1}/\mathcal{B}_{q+1}$. In fact this factor group can best be interpreted as the group "Ext" of group extensions of G by $\mathcal{H}_{q+1}$. The fundamental theorem then has the form

$$H^q(G) = \text{Hom}\,\{\mathcal{H}_q, G\} \times \text{Ext}\,\{G, \mathcal{H}_{q+1}\}.$$

The paper is self contained as far as possible, both in algebraic and topological respects. The first four chapters below develop the requisite group-theoretical notions. Chapter I discusses the groups of homomorphisms involved in the above formula, while Chapter II introduces the group of group extensions, and proves the fundamental theorem relating this group to groups of homomorphisms. This fundamental theorem is essentially a formulation of the known fact that a group extension of G by H can be described either by generators of H (and hence by homomorphisms) or by certain "factor sets." Chapter III analyzes the group Ext $\{G, H\}$ for some special cases of G. Chapter IV introduces some additional groups, closely related to Ext, which arise as inverse limit groups in the treatment of homologies of topological spaces.

The last two chapters analyze homology groups. Chapter V treats the case of a complex, and proves the fundamental theorem quoted above, as well as parallel theorems for some of the other homology groups of a complex. Chapter

[3] More precisely Σ^* is the character group of Σ. The detailed treatment appears in Appendix B below.

VI obtains analogous theorems for the Čech homology groups of a topological space.

Appendix A discusses the case when G is a group with operators. Appendix B contains a computation of the group Ext $\{I, \Sigma^*\}$ mentioned above.

Each chapter is preceded by a brief outline. The chapters are related as in the following diagram:

$$\begin{array}{ccccccc} \mathrm{I} \to \mathrm{II} \to \mathrm{III} & \to & \mathrm{V} \\ & \searrow & & \searrow \\ & & \mathrm{IV} & \to & \mathrm{VI} \end{array}$$

Almost all of V can be read directly after I and II, and a major portion after I alone.

Chapters V and VI are strongly influenced by S. Lefschetz's recent book "Algebraic Topology" [7], that the authors had the privilege of reading in manuscript.

Chapter I. Topological Groups and Homomorphisms

After a certain preliminary definitions, this chapter introduces the basic group Hom $\{R, G\}$ of homomorphisms. In the case when R is a subgroup of a free group, we require two subgroups of "extendable" homomorphisms. The topology of these subgroups is investigated when the "coefficient group" G is itself topological.

1. Topological spaces

A set X is called a *space* if there is given a family of subsets of X, called *open* sets, such that

(1.1) *X and the void set are open,*

(1.2) *the union of any number of open sets is open,*

(1.3) *the intersection of two open sets is open.*

Complements of open sets are called *closed*. X is called a *Hausdorff space* if in addition

(1.4) *every two distinct points are contained respectively in two disjoint open sets.*

X is called a *compact* (= bicompact) space if

(1.5) *every covering of X by open sets contains a finite subcovering.*

A space X is *discrete* if every set in X is open.

The intersection of an open set of a space X with a subset A of X will be called *open in A*. With this convention A becomes a space.

Let X and Y be spaces and $x \to f(x) = y$ a mapping of X into a subset of Y. The mapping f is *continuous* if for every open set $U \subset Y$ the set $f^{-1}(U)$ is open (in X). The mapping f is *open* if for every open set $U \subset X$ the set $f(U)$ is open (in Y). A well known result is

Lemma 1.1. *If f is a continuous mapping of a compact space X into a Hausdorff space Y, then $f(X)$ is closed in Y.*

A *product space* $\prod_\alpha X_\alpha$ of a given collection $\{X_\alpha\}$ of spaces X_α is defined as

the space whose points are all collections $\{x_\alpha\}$, $x_\alpha \in X_\alpha$ and in which open sets are unions of sets of the form $\prod_\alpha U_\alpha$, where U_α is an open subset of X_α and $U_\alpha = X_\alpha$ except for a finite number of indices α.[4] It is known that $\prod X_\alpha$ is a Hausdorff or compact space if and only if for every α the space X_α is a Hausdorff or compact space.[5]

Let Λ be a set of elements and X be a space. We consider the set X^Λ of all functions with arguments in Λ and values in X. The set X^Λ is clearly in a 1-1 correspondence with the product $\prod X_\lambda$ where $\lambda \in \Lambda$ and $X_\lambda = X$. Hence we may consider X^Λ as a space.

2. Topological groups

Only abelian groups (written additively) will be considered.

A group G will be called a *generalized topological group* if G is a space in which the group composition (as a mapping $G \times G \to G$) and the group inverse (as a mapping $G \to G$) are continuous.

If G, considered as a space, is a Hausdorff space, then G will be called a *topological group.*[6] Similarly, if G is compact as a space we shall say that G is a *compact group.*

A subgroup of a (generalized) topological group is a (generalized) topological group. A closed subgroup of a compact group is compact.

Lemma 2.1. *In a generalized topological group G the following properties are equivalent*:

(a) *every point of G is a closed set,*

(b) *the zero element of G is a closed set,*

(c) *G is a topological group.*[7]

The *factor group* $H = G/G_1$ of a generalized topological group G modulo a subgroup G_1 is the group of all cosets $g + G_1$ of G_1 in G. The correspondence $\varphi(G) = H$ carrying each $g \in G$ into its coset $\varphi g = g + G_1$ in H is the "natural" mapping of G on H. We introduce a topology in H by calling a set $U \subset H$ open if and only if $\varphi^{-1}(U)$ is open in G. It can be shown that this topology is the only one under which φ will be both open and continuous.

Lemma 2.2. *If G is a generalized topological group and G_1 is an arbitrary subgroup of G, then the factor group $H = G/G_1$ is a generalized topological group; it is a topological group if and only if G_1 is a closed subgroup of G. If G is compact, then G/G_1 is compact.*

Lemma 2.3. *The closure $\bar{0}$ of the zero element of a generalized topological group is a closed subgroup of G. Its factor group $G/\bar{0}$ is the "largest" factor group of G which is a topological group.*

The preceding two statements show the utility of the study of generalized

---

[4] If $\{\alpha\} = 1, 2, \cdots, n$ we also use the symbol $X_1 \times X_2 \times \cdots \times X_n$ for the product space.

[5] See C. Chevalley and O. Frink, Bulletin Amer. Math. Soc. 47 (1941), pp. 612–614.

[6] G is then a topological group in the sense of Pontrjagin [8].

[7] To prove that a) implies c) one first proves that each neighborhood of g contains the closure of a neighborhood of g, as in Pontrjagin [8], p. 43, proposition F.

topological groups. Several times in the sequel we need to consider an isomorphism

$$G_1/H_1 \cong G_2/H_2 \tag{2.1}$$

where the G_i are topological groups, while the H_i are not closed, so that G_i/H_i are only generalized topological groups. However, if we are able to prove that the isomorphism (2.1) is continuous in both directions in the "generalized" topology of the groups G_i/H_i, we obtain as a corollary the bicontinuous isomorphism of the topological groups $G_i/\bar{H}_i$.

If $\{G_\alpha\}$ is a collection of generalized topological groups the direct product $\prod_\alpha G_\alpha$ is a generalized topological group, provided we define the sum $\{g_\alpha\} = \{g'_\alpha\} + \{g''_\alpha\}$ by setting $g_\alpha = g'_\alpha + g''_\alpha$ for every α. Similarly, if Λ is any set and G is a generalized topological group, then the set G^Λ of all mappings of Λ into G is a generalized topological group. It follows from the results quoted in §1 that $\prod_\alpha G_\alpha$ and G^Λ are topological or compact groups if and only if the groups G_α and G are all topological or compact, respectively.

3. The group of homomorphisms

Let G and H be generalized topological groups. A *homomorphism* θ of H into G is a continuous function $\theta(h)$ defined for all $h \in H$ with values in G, such that $\theta(h_1 + h_2) = \theta(h_1) + \theta(h_2)$. For instance, the natural mapping of a group into one of its factor groups is a homomorphism. If θ_1 and θ_2 are two homomorphisms their sum $\theta_1 + \theta_2$, defined by

$$(\theta_1 + \theta_2)(h) = \theta_1(h) + \theta_2(h), \qquad \text{(all } h \text{ in } H)$$

is also a homomorphism. Under this addition, the set of all homomorphisms θ of H into G constitutes a group, which we denote by Hom $\{H, G\}$:

$$\text{Hom } \{H, G\} = [\text{all homomorphisms } \theta \text{ of } H \text{ into } G]. \tag{3.1}$$

To introduce a (generalized) topology in Hom $\{H, G\}$, take any compact subset X of H and any open subset V of G with $0 \in V$ and consider the set $U(X, V)$ of all θ with $\theta(X) \subset V$. In the usual sense ([8], p. 55) these sets $U(X, V)$ constitute a complete set of neighborhoods of 0 in Hom $\{H, G\}$, and are used to define the topology of Hom $\{H, G\}$.[8]

If H is discrete, the compact subsets X of H are just the finite ones. In this case Hom $\{H, G\}$ is a subgroup of the group G^H with the topology as defined in §2.

LEMMA 3.1. *If G is a topological group and H is discrete, then Hom $\{H, G\}$ is a closed subgroup of the group G^H of all mappings of H into G.*

PROOF. Let $\phi_0 \in G^H$ be a mapping of H into G that is not a homomorphism. There are then elements h_1, h_2, h_3 in H such that $h_1 + h_2 = h_3$ and $\phi_0(h_1) + \phi_0(h_2) \neq \phi_0(h_3)$. Since G is a Hausdorff space and the group composi-

[8] This is the general definition stated by Weil [11], p. 99, and Lefschetz [7], Ch. II.

tion is continuous there are in G three open sets U_1, U_2, U_3 containing $\phi_0(h_1)$, $\phi_0(h_2)$, and $\phi_0(h_3)$, respectively, such that[9] $(U_1 + U_2) \cap U_3 = 0$. Consequently the open subset U of G^H consisting of the mappings ϕ such that $\phi(h_1) \in U_1$, $\phi(h_2) \in U_2$, and $\phi(h_3) \in U_3$ has no elements in common with Hom $\{H, G\}$. Hence Hom $\{H, G\}$ is closed.

COROLLARY 3.2. *If H is discrete and G is a topological (and compact) group, then* Hom $\{H, G\}$ *is a topological (and compact) group.*

Note that the topology of Hom $\{H, G\}$ may not be discrete even though H and G both have discrete topologies. Observe also that if H is discrete, an alteration in the topology of G may alter the topology of Hom $\{H, G\}$ but not its algebraic structure. However, if H carries a non-discrete topology, an alteration in the topology of either H or G may alter the algebraic structure of Hom $\{H, G\}$, in that continuous homomorphisms may cease to be continuous, or vice versa.

If H is compact, we can take H itself to be the compact set X used in the definition of the topology in Hom $\{H, G\}$. Consequently, given any open set V in G containing 0, the homomorphisms θ, such that $\theta(H) \subset V$, constitute an open set. Hence if V can be picked so as not to contain any subgroups but 0, we see that Hom $\{H, G\}$ is discrete.

Subgroups and factor groups of H will correspond respectively to factor groups and subgroups of Hom $\{H, G\}$, as stated in the following lemmas.

LEMMA 3.3. *If H/H_1 is a factor group of the discrete group H, then* Hom $\{H/H_1, G\}$ *is (bicontinuously) isomorphic to that subgroup of* Hom $\{H, G\}$ *which consists of the homomorphisms θ mapping every element of H_1 into zero.*

The proof is readily given by observing that each homomorphism θ with $\theta(H_1) = 0$ maps each coset of H_1 into a single element of G, so induces a homomorphism θ' of H/H_1. The continuity of the isomorphism $\theta \to \theta'$ can be established, as always for isomorphisms between groups, by showing continuity at $\theta = 0$. ([8], p. 63).

LEMMA 3.4. *If L is a subgroup of H, then each homomorphism θ of H into G induces a homomorphism $\theta' = \theta \mid L$ of L into G. The correspondence $\theta \to \theta'$ is a (continuous) homomorphism of* Hom $\{H, G\}$ *into* Hom $\{L, G\}$. *If L is a direct factor of H, this correspondence maps* Hom $\{H, G\}$ *onto* Hom $\{L, G\}$.

4. Free groups and their factor groups

The homology groups will be interpreted later as certain groups of homomorphisms of "free" groups, which we now define. If the elements z_α of a discrete group F are such that every element of F can be represented uniquely as a finite sum $\sum n_\alpha z_\alpha$ with integral coefficients n_α, F is said to be a *free abelian group* with generators (or basis elements) $\{z_\alpha\}$. The number of generators may be infinite. A free group can be constructed with any assigned set of symbols as basis elements.

[9] $U_1 + U_2$ is the set of all sums $g_1 + g_2$, with $g_i \in U_i$. The symbol $\cap$ stands for the set-theoretic intersection.

LEMMA 4.1. *Every proper subgroup of a free group is free.*

For the denumerable case, this is proved by Čech [3]; a general proof is given in Lefschetz [7] (II, (10.1)).

Any discrete group H can be represented as a homomorphic image of a free group. Specifically, if we choose any set of elements t_α in H which together generate all of H, and if we then construct a free group F with generators z_α in 1-1 correspondence $z_\alpha \leftrightarrow t_\alpha$ with the given t's, the correspondence $\sum n_\alpha z_\alpha \to \sum n_\alpha t_\alpha$ will map the free group F homomorphically onto the given group H. If the kernel of this homomorphism[10] is R, H may be represented as the factor group $H = F/R$. R is essentially the group of "relations" on the generators t_α of H.

Given $R \subset F$, each homomorphism ϕ of F into G induces a homomorphism $\theta = \phi \mid R$ of the subgroup R into G, and the homomorphisms so induced form a subgroup of Hom $\{R, G\}$, denoted as

$$\text{(4.1)} \qquad \operatorname{Hom}\{F \mid R, G\} = [\text{all } \theta = \phi \mid R, \text{ for } \phi \in \operatorname{Hom}\{F, G\}].$$

Alternatively, the elements of this subgroup can be described as those homomorphisms θ of R into G which can be extended (in at least one way) to homomorphisms of F into G.

A similar, but lighter, restriction may be imposed as follows: Given $\theta \in \operatorname{Hom}\{R, G\}$, require that for every subgroup $F_0 \supset R$ of F for which F_0/R is finite there exist an extension of θ to a homomorphism of F_0 into G. The θ's meeting this requirement also constitute a subgroup,

$$\text{(4.2)} \qquad \operatorname{Hom}_f\{R, G; F\} = [\text{all } \theta \in \operatorname{Hom}\{F_0 \mid R, G\} \text{ for every finite } F_0/R].$$

These two subgroups,

$$\operatorname{Hom}\{F \mid R, G\} \subset \operatorname{Hom}_f\{R, G; F\} \subset \operatorname{Hom}\{R, G\},$$

are important because the corresponding factor groups in Hom $\{R, G\}$ are invariants of the group $H = F/R$, in that they do not depend on the particular free group F chosen to represent H. This fact may be stated as follows.

THEOREM 4.2. *If H is isomorphic to two factor groups F/R and F'/R' of free groups F and F', then*

$$\text{(4.3)} \qquad \operatorname{Hom}\{R, G\}/\operatorname{Hom}\{F \mid R, G\} \cong \operatorname{Hom}\{R', G\}/\operatorname{Hom}\{F' \mid R', G\},$$

the isomorphism being both algebraic and topological. The same result holds for the factor groups

$$\text{(4.4)} \qquad \operatorname{Hom}\{R, G\}/\operatorname{Hom}_f\{R, G; F\}, \qquad \operatorname{Hom}_f\{R, G; F\}/\operatorname{Hom}\{F \mid R, G\}.$$

This theorem is a corollary of a result to be established in Chapter II, as Theorem 10.1. It can also be proved directly, by appeal to the following lemma, which we state without proof.

---

[10] The *kernel* of a homomorphism θ of a group H is the set of all elements $h \in H$ with $\theta(h) = 0$.

LEMMA 4.3. *Let $F/R = E/G$, where $F \supset R$ is a free group and $E \supset G$ is any other group. There exists a homomorphism ϕ of F into E such that, in the given identification of cosets of G with cosets of R,*

$$\phi(x) + G = x + R, \qquad \textit{for all } x \in F. \tag{4.5}$$

Any other $\phi^* \in$ Hom $\{F, E\}$ with this property (4.5) has the form $\phi^* = \phi + \beta$, for some $\beta \in$ Hom $\{F, G\}$. Conversely, given ϕ with the property (4.5) any such $\phi^* = \phi + \beta$ has the same property.

Although a given group H can be represented in many ways as a factor group $H = F/R$ of a gree group, there is a "natural" such representation, in which F is the additive group F_H of the (integral) group ring of H. Specifically, given H, we choose for each $h \in H$ a symbol z_h and construct a free group F_H generated by the symbols z_h. The correspondence $z_h \to h$ induces a homomorphism of F_H on H. Let R_H denote the kernel of this homomorphism. The factor group (4.3) of the Theorem can then be described invariantly in terms of H and G as the group

$$\text{Hom }\{R_H, G\}/\text{Hom }\{F_H \mid R_H, G\}.$$

The same remark applies to the factor groups of (4.4). It would be possible to use the groups so described as substitutes for the group of group extensions to be introduced in Chapter II.

5. Closures and extendable homomorphisms

If G is topological, we wish to examine the closures of the groups Hom $\{F \mid R, G\}$ and Hom_f in the topological group Hom $\{R, G\}$. A preliminary is a characterization of the subgroup Hom_f.

LEMMA 5.1. *A homomorphism θ of* Hom $\{R, G\}$ *lies in* Hom_f $\{R, G; F\}$ *if and only if for each element t in F with a multiple mt in R there exists $h \in G$ with $\theta(mt) = mh$.*

PROOF. Let F_t be the subgroup of F generated by t and R. If $mt \in R$ for $m \neq 0$, F_t/R is finite and cyclic, so that $\theta \in \text{Hom}_f$ is extendable to F_t. Hence the condition stated on $\theta(mt)$ is necessary. Conversely, for any given group $F_0 \subset F$ with F_0/R finite we can write F_0/R as a direct product of cyclic groups. By applying the given condition on θ to each of these cyclic groups, we find an extension of θ to F_0, as required.

Another characterization of Hom_f can be found; the proof is similar:

LEMMA 5.2. *A homomorphism θ of* Hom $\{R, G\}$ *lies in* Hom_f $\{R, G; F\}$ *if and only if θ can be extended to a homomorphism (into G) of each subgroup F_0 of F which contains R and for which the factor group F_0/R has a finite number of generators.*

We now consider the topology on Hom $\{R, G\}$.

LEMMA 5.3. *If G and hence* Hom $\{R, G\}$ *are generalized topological groups,* Hom_f $\{R, G; F\}$ *is contained in the closure of* Hom $\{F \mid R, G\}$, or

$$\mathrm{Hom}\,\{F \mid R, G\} \subset \mathrm{Hom}_f\,\{R, G; F\} \subset \overline{\mathrm{Hom}}\,\{F \mid R, G\} \subset \mathrm{Hom}\,\{R, G\}.$$

PROOF. Let θ_0 be in $\mathrm{Hom}_f\,\{R, G; F\}$, while U is any open set of $\mathrm{Hom}\,\{R, G\}$ containing θ_0. Since F is discrete, the definition of the topology in $\mathrm{Hom}\,\{R, G\}$ implies that there is a finite set of elements $r_1, \cdots, r_n$ of R such that U contains all θ for which each $\theta(r_i) = \theta_0(r_i)$. The elements r_i are all contained in a subgroup F_0 of F generated by a finite number of the given independent generators of the free group F. Since $\theta_0 \in \mathrm{Hom}_f$, θ_0 has an extension θ' to the group generated by F_0 and R (Lemma 5.2). Introduce a new homomorphism θ^* of F by setting $\theta^*(z_\alpha) = \theta'(z_\alpha)$ for each generator z_α of F_0, $\theta^*(z_\alpha) = 0$ otherwise. This θ^* induces a homomorphism θ of R, which agrees with θ_0 on the original elements $r_1, \cdots, r_n$ and which is by construction an element of $\mathrm{Hom}\,\{F \mid R, G\}$. In other words, the arbitrary neighborhood U of θ_0 does contain a homomorphism $\theta \in \mathrm{Hom}\,\{F \mid R, G\}$. This proves the lemma.

LEMMA 5.4. *If G is a compact topological group,* $\mathrm{Hom}\,\{F \mid R, G\}$ *is a closed sub-group of* $\mathrm{Hom}\,\{R, G\}$, *and hence* $\mathrm{Hom}\,\{F \mid R, G\} = \mathrm{Hom}_f\,\{R, G; F\}$.

PROOF. By Corollary 3.2, both the groups $\mathrm{Hom}\,\{R, G\}$ and $\mathrm{Hom}\,\{F, G\}$ are compact and topological. The second of these groups is mapped homomorphically onto $\mathrm{Hom}\,\{F \mid R, G\}$ by the continuous correspondence $\theta \to \theta \mid R$ of Lemma 3.4. Therefore, by Lemma 1.1, the image $\mathrm{Hom}\,\{F \mid R, G\}$ is closed.

For any integer m, let mG be the subgroup of all elements of the form mg, with g in G. A condition for the closure of Hom_f may be stated in terms of these subgroups.

LEMMA 5.5. *If G is a generalized topological group, then* $\mathrm{Hom}_f\,\{R, G; F\}$ *is closed in* $\mathrm{Hom}\,\{R, G\}$ *whenever every subgroup mG of G is closed in G, for $m = 2, 3, \cdots$.*[11]

PROOF. Let θ be a homomorphism in the closure of $\mathrm{Hom}_f\,\{R, G; F\}$. Consider an arbitrary t in F such that $mt \in R$. By Lemma 5.1 and the given condition on G it will suffice to prove that $\theta(mt) \in \overline{mG}$. Let V be any open set containing 0 in G. By the definition of the topology in $\mathrm{Hom}\,\{R, G\}$, there exists for θ in the closure of Hom_f an element θ' in Hom_f itself, such that $\theta'(mt) - \theta(mt) \in V$. But $\theta'(mt)$ is in mG, so that the arbitrary open set $V + \theta(mt)$ does contain an element of mG. This proves $\theta(mt)$ in $\overline{mG}$, as required.

An examination of this proof shows that the given condition on G can be somewhat weakened. It suffices to require that the subgroup mG be closed in G for every integer m which is the order of an element of F/R. The same remark will apply in various subsequent cases when this condition on G is used.

CHAPTER II. GROUP EXTENSIONS

This chapter introduces the basic group $\mathrm{Ext}\,\{G, H\}$ of all group extensions of G by H, and its subgroup $\mathrm{Ext}_f\,\{G, H\}$ of all extensions which are "finitely trivial"

[11] If every subgroup mG is closed in G, Steenrod [9] and Lefschetz [7] say that G has the "division closure property."

(§8). Each individual group extension can be described either by a suitable "factor set" (§7) or by a certain homomorphism. The equivalence of these two representations is the fundamental theorem of this chapter (Theorem 10.1); it gives an expression of Ext $\{G, H\}$ as one of the factor-homomorphism groups already considered in Chapter I. This fundamental theorem, which is implicit in previous algebraic work on group extensions, is of independent algebraic interest. The chapter closes with a proof that the representation of Ext $\{G, H\}$ by homomorphisms is a "natural" one (§12). This conclusion is needed for the subsequent limiting process, which is used in defining the Čech homology groups.

6. Definition of extensions

A group E having G as subgroup and $H = E/G$ as the corresponding factor group is said to be an "extension" of G by H. More explicitly, if the groups G and H are given, a *group extension* of G by H is a pair (E, β), where E is a group containing G and β is a homomorphism of E onto H under which exactly the elements of G are mapped into $0 \in H$.[12] Such a β induces an isomorphism of E/G to H. For given G and H, two extensions (E_1, β_1) and (E_2, β_2) are regarded as *equivalent* if and only if there is an isomorphism ω of E_1 to E_2 which leaves elements of G and cosets of H fixed. In other words, the isomorphism ω of E_1 to E_2 must have $\omega g = g$ for $g \in G$ and $\beta_2 \omega x = \beta_1 x$ for $x \in E_1$. We regard equivalent extensions as identical, and so study the equivalence classes of extensions of G by H. It will appear that these equivalence classes are themselves the elements of a group.

For given G and H, the direct product $G \times H$ has the "natural" homomorphism $(g, h) \rightarrow h$ onto H, and so can be regarded as an extension of G by H. Any extension (E, β) equivalent to this direct product (with its natural homomorphism) is said to be a *trivial* extension of G by H.

7. Factor sets for extensions

A given extension (E, β) of G by H can be described in terms of representatives for elements of H. To each h in H select in E a representative $u(h)$, such that $\beta(u(h)) = h$. Every element of E lies in some coset h, so has the form $g + u(h)$ for g in G. The sum of any two representatives $u(h)$ and $u(k)$ will lie in the same coset, modulo G, as does the representative of the sum $h + k$. Hence there is an addition table of the form

$$u(h) + u(k) = u(h + k) + f(h, k), \tag{7.1}$$

where $f(h, k)$ lies in G for each pair of elements h, k in H. The commutative and associative laws in the group E imply two corresponding identities for f,

$$f(h, k) = f(k, h), \tag{7.2}$$

[12] Group extensions are discussed by Baer [2], Hall [6], Turing [11], Zassenhaus [15], and elsewhere. Much of the discussion in the literature treats the more general case in which G but not H is assumed to be abelian and in which G is not necessarily in the center of H.

$$f(h, k) + f(h + k, l) = f(h, k + l) + f(k, l). \tag{7.3}$$

The sum of any two elements $g_1 + u(h)$ and $g_2 + u(k)$ of E is determined by the addition table (7.1) and the addition given within G and H.

The extension E does not uniquely determine the corresponding function f. An arbitrary set of representatives $u'(h)$ for the elements of H can be expressed in terms of the given representatives as

$$u'(h) = u(h) + g(h), \qquad \text{each } g(h) \in G;$$

they will have an addition table like that of (7.1) with a function f' given by

$$f'(h, k) = f(h, k) + [g(h) + g(k) - g(h + k)]. \tag{7.4}$$

Conversely, a *factor set* of H in G is any function $f(h, k)$, with values in G for h, k in H which satisfies the "commutative" and "associative" conditions (7.2) and (7.3) for all h, k, and l in H. A *transformation set* is any function of h and k like the term in brackets in (7.4); thus for any function $g(h)$ defined for each $h \in H$ and taking on values in G, the function

$$t(h, k) = g(h) + g(k) - g(h + k) \tag{7.5}$$

is a transformation set. Such a set automatically satisfies the conditions (7.2) and (7.3), hence is always a factor set. Two factor sets f and f' are said to be *associate* if their difference is, as in (7.4), a transformation set. The correspondence between group extensions and factor sets may now be formulated as follows.

THEOREM 7.1. *For given groups G and H, there is a many-one correspondence $f \to (E, \beta)$ between the factor sets f of H in G and the group extensions (E, β) of G by H, where $f \to (E, \beta)$ holds if and only if f is the factor set which appears in one of the possible "addition tables" (7.1) for E. Two factor sets f and f' of H in G determine equivalent group extensions of G by H if and only if they are associate. In particular, the group extension determined by f is trivial if and only if f is a transformation set.*

PROOF. As a preliminary, observe that the associative relations (7.3) for f show (with $k = l = 0$, $h = k = 0$) that $f(0, 0) = f(h, 0) = f(0, l)$. Now, given f, we construct E_f as the group of all pairs (g, h) with addition given by the rule

$$(g_1, h) + (g_2, k) = (g_1 + g_2 + f(h, k), h + k),$$

and the homomorphism β_f defined by $\beta_f(g, h) = h$. Since $f(0, 0) = f(0, l)$, each element $(g, 0)$ may be identified with the corresponding element $g + f(0, 0)$ in G; the pair (E_f, β_f) is then indeed an extension of G by H. As a representative of h in E_f, we may choose $u(h) = (0, h)$; the addition table (7.1) then involves exactly the original factor set f. If E is an arbitrary group extension

of G by H in which f appears as the factor set of E, the correspondence $g + u(h) \leftrightarrow (g, h)$ shows that the extension E is in fact equivalent to the extension E_f just constructed. Therefore $f \rightarrow (E_f, \beta_f)$ is a many-one correspondence with the defining property stated in the theorem.

If f and f' are associate, as in (7.4), the correspondence

$$(g, h) \rightarrow (g - g(h), h)'$$

shows that the corresponding extensions E_f and $E_{f'}$ are equivalent. Conversely, the argument leading to (7.4) shows in effect that E_f is equivalent to $E_{f'}$ only if f is associate to f'.

We turn now to two special applications of transformation sets. In the first place, the representative for the zero element of H may always be chosen as the zero in E. This means that $u'(0) = 0$, $u'(0) + u'(h) = u'(h)$, so that

$$f'(0, h) = f'(h, 0) = 0 \qquad \text{(all } h \in H). \tag{7.6}$$

A factor set f' with the property (7.6) may be called *normalized*; we have proved that every factor set f is associate to a normalized factor set.

Free groups may be characterized in terms of group extensions as follows:

THEOREM 7.2. *A group with more than one element H is free if and only if every extension of any group by H is the trivial extension.*

PROOF. Suppose first that H satisfies the condition that every extension of every G is trivial. Represent H as F/R, where F is free. Then F is a trivial extension of R by H, hence is a direct sum of R and H. Therefore H, as a subgroup of the free group F, is itself free. The other half of the theorem is stated in more detail in the following Lemma.

LEMMA 7.3. *Every factor set f' of a free group F in a group G is a transformation set, so that*

$$f'(x, y) = \phi(x + y) - \phi(x) - \phi(y), \qquad \phi(x) \in G, \tag{7.7}$$

holds for all $x, y \in F$. If F has generators z_α, the function ϕ may be chosen so that $\phi(0) = -f'(0, 0)$, $\phi(z_\alpha) = 0$ for each generator z_α.

PROOF. In the extension $E_{f'}$ of G by F we have an addition table

$$u'(x) + u'(y) = u'(x + y) + f'(x, y) \qquad (x, y \in F).$$

In E we introduce a new set of representatives $u(\sum e_\alpha z_\alpha) = \sum e_\alpha u'(z_\alpha)$ for the elements $\sum e_\alpha z_\alpha$ of F. These are related to the original representatives by an equation $u(z) = u'(z) + \phi(z)$, where $\phi(z)$ has values in G. Because F is a free group, $z \rightarrow u(z)$ as defined is a homomorphism of F into E, so that $u(x + y) = u(x) + u(y)$, and the factor set belonging to u is identically zero. But the given f' is associate to this zero factor set, as in (7.4). Setting $f = 0$, $\phi = -g$ in (7.4) gives (7.7), as desired. By construction, $u(z_\alpha) = u'(z_\alpha)$, so $\phi(z_\alpha) = 0$. Also $u'(0) + u'(0) = u'(0) + f'(0, 0)$, so that $u'(0) = f'(0, 0)$, $u(0) = 0$, and therefore $\phi(0) = -f'(0, 0)$. This completes the proof.

8. The group of extensions

For fixed H and G the sum of two factor sets f_1 and f_2 is a third factor set, defined as

$$(f_1 + f_2)(h, k) = f_1(h, k) + f_2(h, k) \qquad (h, k \in H).$$

Under this addition, the factor sets and the transformation sets form groups, denoted respectively by

(8.1) $\quad$ Fact $\{G, H\}$ = group of all factor sets of H in G,

(8.2) $\quad$ Trans $\{G, H\}$ = group of all transformation sets of H in G.

The factor sets belonging to a given group extension E constitute a coset of the subgroup Trans $\{G, H\}$, as in (7.4). Hence the correspondence of factor sets to extensions is a one-one correspondence between cosets of Fact/Trans and equivalence classes of extensions. This correspondence carries the addition of factor sets into an addition of group extensions. We are thus led to define the *group of group extensions* of G by H as[13]

(8.3) $\quad$ Ext $\{G, H\}$ = Fact $\{G, H\}$/Trans $\{G, H\}$.

If H is discrete while G is a (generalized) topological group, there will be a corresponding induced topology on Ext $\{G, H\}$. For each factor set f is a function on $H \times H$ with values in G, so that Fact $\{G, H\}$ is a subgroup of the generalized topological group $G^{H \times H}$ of all such functions. The subgroup "Trans" and the factor group "Ext" also carry topologies. Much as in §3 one can prove that if H is discrete and G topological, then Fact $\{G, H\}$ is a closed subgroup of $G^{H \times H}$. This proves

LEMMA 8.1. *If H is discrete and G is a topological (and compact) group, then* Fact $\{G, H\}$ *is a topological (and compact) group.*

In general, however, Trans $\{G, H\}$ will not be closed in Fact $\{G, H\}$, even when G is topological. In such cases Ext $\{G, H\}$ is necessarily a generalized topological group.

If (E, β) is an extension of G by H, each subgroup $S \subset H$ determines a corresponding subgroup $E_S \subset E$, consisting of all $e \in E$ with $\beta(e) \in S$. Since $E_S \supset G$, we may thus say that E "induces" an extension (E_S, β) of G by S. We call an extension E *finitely trivial* if E_S is trivial for every finite subgroup $S \subset H$.

Similarly, any factor set f of H in G determines for each subgroup $S \subset H$ a factor set f_S of S in G, where $f_S(h, k) = f(h, k)$ for h, k in S (i.e., f_S is obtained by "cutting off" f at S). The correspondence between factor sets and group extensions readily gives

LEMMA 8.2. *A factor set f of H in G determines a finitely trivial extension of*

[13] It is possible to define the sum of two group extensions directly, without using the factor sets (see Baer [2] p. 394); it also is possible to give an analogous definition of the topology introduced below in Ext $\{G, H\}$.

G by H if and only if, for every finite subgroup $S \subset H$, the factor set f_S "cut off" at S is a transformation set of S in G. Hence the finitely trivial extensions of G by H constitute a subgroup $\mathrm{Ext}_f \{G, H\}$ *of* $\mathrm{Ext}\ \{G, H\}$.

9. Group extensions and generators

A group extension can be described not only by factor sets, but also by certain homomorphisms related to the generators of the extending group H. For let (E, β) be a given extension of G by H, and $H = F/R$ a representation of H as a factor group of a free group F. Let F have the generators z_α, as in §4; the corresponding elements (or cosets) t_α of H will then be a set of generators of H. For each generator t_α choose a corresponding representative u_α in the given group extension E, so that $\beta u_\alpha = t_\alpha$. Then $\beta(\sum e_\alpha u_\alpha) = \sum e_\alpha t_\alpha$, so that any element $\sum e_\alpha t_\alpha \in H$ has a representative of the form $\sum e_\alpha u_\alpha$. This means that each element of E can be written in the form

$$x = g + \sum e_\alpha u_\alpha, \qquad g \in G, \qquad e_\alpha \text{ integers}.$$

From this representation one can at once determine how to add the elements of E. However, this representation is not in general unique, for $(\sum e_\alpha u_\alpha) \in G$ is equivalent to $\sum e_\alpha t_\alpha = 0$, which in turn is equivalent to $(\sum e_\alpha z_\alpha) \in R$. Thus to each $r = \sum e_\alpha z_\alpha$ in the group R of "relations" there is assigned an element $\theta(r) \in G$, defined as

$$\theta(r) = \theta(\sum e_\alpha z_\alpha) = \sum e_\alpha u_\alpha$$

These assignments $\theta(r)$ completely determine the extension E.

The function θ hereby defined[14] is a homomorphism of R into G. Conversely every such homomorphism θ can be used to construct a corresponding group extension of G by $H = F/R$; it suffices to construct E by reducing the direct product $F \times G$ modulo the subgroup of all elements of the form $(r, \theta(r))$, for $r \in R$. There is thus a correspondence between homomorphisms of R into G and extensions of G by $H = F/R$.[15]

10. The connection between homomorphisms and factor sets

Given G and $H = F/R$, an extension E of G by H may be given either by a factor set or by a homomorphism of R into G. There must therefore be a relation between factor sets and homomorphisms of this type. We now propose to establish this relation directly, without using extensions explicitly. (Actually, the correspondence which we obtain is identical with that obtained by going from a homomorphism first to the corresponding group extension and then to its factor set.)

THEOREM 10.1. *If $H = F/R$ is a factor group of a free group F, while G is any other group, then*

---

[14] Actually, θ may be obtained by "cutting off" one of the homomorphisms ϕ as described in Lemma 4.3.

[15] This correspondence has been stated by Baer ([2], p. 395) and used by Hall [6].

(10.1) $$\mathrm{Ext}\,\{G, H\} \cong \mathrm{Hom}\,\{R, G\}/\mathrm{Hom}\,\{F \mid R, G\}.$$

Under the correspondence which gives this isomorphism

(10.2) $$\mathrm{Ext}_f\,\{G, H\} \cong \mathrm{Hom}_f\,\{R, G; F\}/\mathrm{Hom}\,\{F \mid R, G\},$$

(10.3) $$\mathrm{Ext}\,\{G, H\}/\mathrm{Ext}_f\,\{G, H\} \cong \mathrm{Hom}\,\{R, G\}/\mathrm{Hom}_f\,\{R, G; F\}.$$

If G is a generalized topological group while F and H are discrete, all these isomorphisms are bicontinuous.

PROOF. As a preliminary, observe that the representation $H = F/R$ means that the free group F is a group extension of R by H. In this extension choose a representative $u_0(h)$ in F for each $h \in H$. F is then described, as in (7.1), by an addition table

(10.4) $$u_0(h) + u_0(k) = u_0(h + k) + f_0(h, k),$$

where f_0 is a factor set of H in R. This factor set will be fixed throughout the proof.

Since $\mathrm{Ext}\,\{G, H\}$ is defined as Fact/Trans, the required isomorphism (10.1) could be established by a suitable correspondence of homomorphisms to factor sets. Let $\theta \in \mathrm{Hom}\,\{R, G\}$ be given, and define f_θ by

(10.5) $$f_\theta(h, k) = \theta[f_0(h, k)] \qquad (h, k \in H).$$

The requisite commutative and associative laws (7.2) and (7.3) for f_θ follow from those for f_0, and the correspondence $\theta \to f_\theta$ is a homomorphism of $\mathrm{Hom}\,\{R, G\}$ into $\mathrm{Fact}\,\{G, H\}$, and therefore into $\mathrm{Ext}\,\{G, H\}$.

Suppose next that θ can be extended to a homomorphism θ^* of F into G. This homomorphism applied to (10.4) gives

$$\theta^*[f_0(h, k)] = \theta^*[u_0(h)] + \theta^*[u_0(k)] - \theta^*[u_0(h + k)].$$

If we set $g(h) = \theta^*[u_0(h)]$, the result asserts that $\theta^* f_0 = \theta f_0 = f_\theta$ is a transformation set.

Conversely, suppose that f_θ is a transformation set, so that $f_\theta(h, k) = g(h) + g(k) - g(h + k)$ for some function g. Now any element in F can be written, in only one way, in the form $r + u_0(h)$, with r in R, h in H. We define $\theta^*(r + u_0(h))$ as $\theta(r) + g(h)$. Clearly θ^* is an extension of θ; a straightforward computation with (10.4) shows that θ^* is actually a homomorphism. In this case, then, θ is extendable to F.

We know now that the correspondence $\theta \to f_\theta$ is an isomorphism of $\mathrm{Hom}\,\{R, G\}/\mathrm{Hom}\,\{F \mid R, G\}$ into a subgroup of $\mathrm{Ext}\,\{G, H\}$. It remains to prove that it is a homomorphism onto. At this juncture we use for the first time the assumption that F is a free group. Let E be a given extension of G by H, with a factor set f which we can assume is normalized, as in (7.6). Let β_0 be the given homomorphism of F on H. Use f to define a factor set f' of F in G by the equation

$$f'(x, y) = f(\beta_0 x, \beta_0 y), \qquad x, y \in F. \tag{10.6}$$

Since F is free, f' is a transformation set, so we can find, as in Lemma 7.3, a function $\phi(z)$ on F to G with

$$\phi(x + y) = \phi(x) + \phi(y) + f'(x, y). \tag{10.7}$$

In particular, if x and y lie in R, $\beta_0 x = \beta_0 y = 0$, and $f'(x, y) = f(0, 0) = 0$, because f is normalized. Thus ϕ, restricted to R, is a homomorphism $\theta = \phi \mid R$ of R into G. Furthermore, if ϕ is applied to the addition table (10.4) for F, the property (10.7) gives

$$\phi[u_0(h) + u_0(k)] = \phi[u_0(h + k)] + \phi[f_0(h, k)],$$

where a term $f'(u_0(h + k), f_0(h, k))$, which would have entered by (10.7), is zero because f is normalized, $f_0(h, k) \in R$, and $\beta_0 f_0(h, k) = 0$. Now compute $f(h, k)$ for h, k in H. By (10.6),

$$\begin{aligned} f(h, k) &= f'(u_0(h), u_0(k)) \\ &= \phi[u_0(h) + u_0(k)] - \phi[u_0(h)] - \phi[u_0(k)] \\ &= \phi[u_0(h + k)] - \phi[u_0(h)] - \phi[u_0(k)] + \phi[f_0(h, k)], \end{aligned}$$

in virtue of the equation displayed just above. This equation asserts that f is associate to the factor set $\phi f_0 = \theta f_0$. In other words, given the normalized factor set f, we have constructed a homomorphism θ for which f is essentially θf_0. This completes the proof of (10.1).

It is desirable to find a more explicit expression for this dependence of θ on f. A simple induction applied to (10.7) will show that, for z_i in F,

$$\phi\left(\sum_{i=1}^{n} z_i\right) = \sum_{i=1}^{n} \phi(z_i) + \sum_{k=1}^{n-1} f'\left(\sum_{i=1}^{k} z_i,\, z_{k+1}\right).$$

If z_i is one of the generators z_α of F, then $\phi(z_i) = 0$, by Lemma 7.2. If $z_i = -z_\alpha$ is the negative of a generator, then by (10.7)

$$\phi(0) = \phi(z_\alpha + (-z_\alpha)) = \phi(z_\alpha) + \phi(-z_\alpha) + f'(z_\alpha, -z_\alpha),$$

so that $\phi(-z_\alpha) = -f'(z_\alpha, -z_\alpha)$. Now any element of F can be written as a finite linear combinations of generators and hence as a sum $\sum x_i$, where each x_i is either a generator or the negative of a generator z_α, and where any given generator may appear several times in this sum. In particular, for any element $r = \sum x_i$ in the subgroup R, the previous formula for ϕ becomes a formula for $\theta = \phi \mid R$,

$$\theta\left(\sum_{i=1}^{n} x_i\right) = -\sum{}' f(\beta_0 x_i, -\beta_0 x_i) + \sum_{k=1}^{n-1} f\left(\sum_{i=1}^{k} \beta_0 x_i,\, \beta_0 x_{k+1}\right), \tag{10.8}$$

where β_0 is the given homomorphism of F into H, and where the sum $\sum'$ is taken over those elements x_i which are the negatives of generators. The

essential feature of this formula is the fact that it expresses $\theta(r)$ for $r \in R$ as a sum of a finite number of values of the given factor set f of H in G.

Now consider the continuity of the correspondence $\theta \to f_\theta$ used to establish (10.1). It suffices to establish the continuity at 0. If U is any open set, containing zero, in Hom $\{R, G\}$/Hom $\{F \mid R, G\}$, there will be an open set V containing 0 in G and a finite set of elements $r_1, \cdots, r_s \in R$ such that U contains the cosets of all homomorphisms θ with $\theta(r_i) \in V$, $i = 1, \cdots, s$.

For a given f, the expressions $\theta(r_i)$ of (10.8) for these elements r will involve but a finite number of elements of the factor set f. Because of the continuity of addition in G, we can construct an open set U' in Fact $\{G, H\}$ such that each $\theta(r_i)$ does in fact lie in the given V. This establishes the continuity of the correspondence $f \to \theta$. The continuity of the inverse correspondence is obtained by a similar argument on the definition (10.5) of this correspondence.

It remains only to consider the formulas (10.2) and (10.3) on finitely trivial extensions. Let θ and its correspondent f_θ be given, and let $F_0 \supset R$ be any subgroup of F for which F_0/R is finite. A previous argument, applied to F_0 instead of F, shows that θ can be extended to a homomorphism of F_0 into G if and only if f_θ, regarded as a factor set for F_0/R in G, is a transformation set. But the subgroup $\text{Hom}_f \{R, G; F\}$ by definition consists of all those θ which are extendable to every such F_0, while Ext_f by Lemma 8.2 is obtained from those factor sets which are transformation sets on every such subgroup F_0. $\text{Hom}_f \{R, G; F\}$/Hom $\{F/R, G\}$ is the subgroup corresponding to $\text{Ext}_f \{G, H\}$ under $\theta \to f_\theta$. This proves (10.2) and with it (10.3). The continuity of the isomorphisms in this case follows from the continuity of the isomorphism (10.1).

For subsequent purposes we observe that the correspondence $\theta \to f_\theta$ obtained in this proof is essentially independent of the choice of the fixed factor set f_0 for H in R. Specifically, if f_0 is replaced by an associate factor set f_0', f_θ will be replaced also by an associate factor set, so that the corresponding element of Ext $\{G, H\}$ is not altered.

11. Applications

The representation of Ext $\{G, H\}$ as Hom $\{R, G\}$/Hom $\{F \mid R, G\}$ gives an immediate proof of the invariance of the latter group, as stated in Theorem 4.2 of Chapter I. There are a number of other simple corollaries.

COROLLARY 11.1. *For a direct product $H \times H'$,*

$$\text{Ext}\,\{G, H \times H'\} \cong \text{Ext}\,\{G, H\} \times \text{Ext}\,\{G, H'\}. \tag{11.1}$$

If G is a generalized topological group, the isomorphism is bicontinuous.

PROOF. If $H = F/R$ and $H' = F'/R'$, we may write $H \times H' = (F \times F')/(R \times R')$, where $F \times F'$, like F and F', is free. Each homomorphism of $R \times R'$ into G determines homomorphisms θ and θ' of the subgroups R and R' into G, and this correspondence yields a (bicontinuous) isomorphism

$$\text{Hom}\,\{R \times R', G\} \cong \text{Hom}\,\{R, G\} \times \text{Hom}\,\{R', G\}.$$

Furthermore, under the same correspondence

$$\text{Hom}\,\{(F \times F') \mid (R \times R'), G\} \cong \text{Hom}\,\{F \mid R, G\} \times \text{Hom}\,\{F' \mid R', G\}.$$

These two relations yield a corresponding isomorphism between the respective factor groups such as Hom $\{R, G\}$/Hom $\{F \mid R, G\}$. By the fundamental theorem, the latter isomorphism is the one asserted in (11.1).

This conclusion can also be established without using homomorphisms, by a direct argument like that of Lemma 7.2. (Choose new representatives in E for elements of $H \times H'$ by setting $u'(hh') = u(h)u(h')$). Another simple argument directly with the factor sets will give a companion "direct product" representation,

$$\text{Ext}\,\{G \times G', H\} \cong \text{Ext}\,\{G, H\} \times \text{Ext}\,\{G', H\}; \tag{11.2}$$

this isomorphism is also bicontinuous.

Corollary 11.2. *If H is a cyclic group of order m, then*

$$\text{Ext}\,\{G, H\} \cong G/mG, \qquad (mG = \textit{all } mg, \textit{for } g \in G). \tag{11.3}$$

This isomorphism is also bicontinuous.

This is a well known result, which can be derived directly from our main theorem. The cyclic group H can be written as $H = F/R$, where F is an infinite cyclic group with generator z, R the subgroup generated by mz. Then any $\theta \in \text{Hom}\,\{R, G\}$ is uniquely determined by the image $\theta(mz) = h$ of the generator mz. This correspondence $\theta \to h$ (mod mG) gives the isomorphism (11.3).

A similar representation can be found for any finite abelian group H, simply by representing H as a direct product of cyclic groups of orders m_i, $i = 1, \cdots, t$. By Corollary 11.1, Ext $\{G, H\}$ is then isomorphic to the direct product of the groups G/m_iG. A similar decomposition applies if the abelian group H has a finite number of generators. The result may be stated as follows.

Corollary 11.3. *If H has a finite number of generators, and T is the subgroup of all elements of finite order in H, then* Ext $\{G, H\} \cong$ Ext $\{G, T\}$, *algebraically and topologically. The latter group is a direct product of groups of the form G/mG.*

Theorem 7.2 (extensions by a free group are trivial) has an analogue for infinitely divisible groups. Recall that G is *infinitely divisible* if for each $g \in G$ and each integer $m \neq 0$ the equation $mx = g$ has a solution $x \in G$.

Corollary 11.4. *A group G is infinitely divisible if and only if every extension of G by any group is the trivial extension.*

Proof. If G is not infinitely divisible, some $G/mG \neq 0$, so that there will be a non-trivial extension of G by a cyclic group, as in Corollary 11.2. Conversely, suppose G is infinitely divisible. If $R \subset F$ are groups, a transfinite induction will show that every homomorphism of R into G can be extended to a homomorphism into G of the larger group F. Therefore the subgroup Hom $\{F \mid R, G\}$ exhausts the group Hom $\{R, G\}$, and Ext $\{G, F/R\} = 0$.

Corollary 11.5. *If T is the subgroup of all elements of finite order in H, then*

(11.4) $\qquad \operatorname{Ext}\{G, H\}/\operatorname{Ext}_f\{G, H\} \cong \operatorname{Ext}\{G, T\}/\operatorname{Ext}_f\{G, T\}.$

This isomorphism is bicontinuous (if G is a generalized topological group).

PROOF. In the representation $H = F/R$, let F_T denote the set of all elements of F of finite order modulo R. The group T then has the representation $T = F_T/R$, while F_T, as a subgroup of a free group, is itself free. Now the group $\operatorname{Hom}_f\{R, G; F\}$ by definition consists of all homomorphisms extendable to subgroups of F finite over R; as these subgroups are all contained in F_T, the group Hom_f is identical with $\operatorname{Hom}_f\{R, G; F_T\}$. If both factor groups in (11.4) are now represented by groups of homomorphisms, as in (10.3), the result is immediate.

Observe that when T has only elements of finite order, the group $\operatorname{Ext}_f\{G, T\}$, though it consists of extensions E of G by T trivial on every finite subgroup of T, can contain non-trivial extensions. This is illustrated by the following example. Let p be a prime, and G a group with generators $g, h_1, h_2, \cdots$ and relations $p^i h_i = g$, for $i = 1, 2, \cdots$. In this group G the intersection of all the subgroups p^iG is the group generated by g alone. Let T be the group of all rational numbers of the form a/p^i, reduced modulo 1. Then all elements of T have finite order, and T may be written as $T = F/R$, where F is a free group with generators $z_1, z_2, \cdots$, and R the free subgroup generated by pz_1, $pz_2 - z_1, pz_3 - z_2, \cdots$. (The homomorphism $F \to T$ maps z_i into $1/p^i$.)

To prove $\operatorname{Ext}_f\{G, T\} \neq 0$ it suffices to find a $\theta \in \operatorname{Hom}_f\{R, G; F\}$ which is not in $\operatorname{Hom}\{F \mid R, G\}$. Such a θ is determined by setting $\theta(pz_1) = g$, $\theta(pz_{i+1} - z_i) = 0$, $i = 1, 2, \cdots$. The definition $\theta^*(z_{n-i}) = p^i h_n$ will provide an extension θ^* of θ to the finite subgroup of F generated by $z_1, \cdots, z_n$. However, suppose that θ had an extension ϕ to F. Then $\phi(pz_{i+1}) = \phi(z_i)$, so that $\phi(z_1) = p^n\phi(z_{n+1})$ for every n. This means that $\phi(z_1)$ is in every subgroup p^nG, hence has the form eg for an integer e. But then $g = \theta(pz_1) = p\phi(z_1) = epg$ gives a contradiction. Therefore $\operatorname{Ext}_f\{G, T\} \neq 0$ in this case. However, if G has no elements of finite order, one can prove easily that $\operatorname{Ext}_f\{G, T\} = 0$, using Lemma 5.1 (see §17 below).

For several types of topological groups G, §5 gives information on the topology of the various relevant subgroups of $\operatorname{Hom}\{R, G\}$. By the main theorem, the conclusions of Lemmas 5.3, 5.4, and 5.5 can now be rewritten as conclusions about the topology of $\operatorname{Ext}\{G, H\}$, as follows.

COROLLARY 11.6. *If H is discrete and G a generalized topological group, the closure of the zero element in the generalized topological group* $\operatorname{Ext}\{G, H\}$ *contains* $\operatorname{Ext}_f\{G, H\}$. *If, in addition, every subgroup mG is closed in G, for $m = 2, 3, \cdots$, then* $\operatorname{Ext}_f\{G, H\}$ *is closed in* $\operatorname{Ext}\{G, H\}$.

In particular, if H has no elements of finite order, then every extension of G by H is trivial on (the non-existent) finite subgroups of H, consequently $\operatorname{Ext}_f\{G, H\} = \operatorname{Ext}\{G, H\}$ and the closure of 0 is the whole group $\operatorname{Ext}\{G, H\}$. This means that $\operatorname{Ext}\{G, H\}$ carries the "trivial" (generalized) topology in which the only open sets are the whole group and the empty set.

Corollary 11.7. *If H is discrete and G compact and topological, then* $\mathrm{Ext}_f\,\{G, H\} = 0$ *and* Ext $\{G, H\}$ *is itself a compact topological group.*

This conclusion is obtained from Lemma 8.1 and from Lemma 5.4.

12. Natural homomorphisms

The basic homomorphism $\eta(\theta) = f_\theta$ mapping elements θ of Hom $\{R, G\}$ into factor sets f, as in Theorem 10.1, is a "natural" one. Specifically, this means that the application of η "commutes" with the application of any homomorphism T to the free group F and its subgroup R. To state this more precisely, we need to consider first the homomorphisms which T induces on the groups Hom $\{R, G\}$ and Ext $\{G, H\}$.

Let F' be a free group with subgroup R', T a homomorphism $z' \to Tz'$ of F' into the free group F such that $T(R') \subset R$. T induces a homomorphism of $H' = F'/R'$ into $H = F/R$. This induced homomorphism will be written with the same letter T, so that $T(g + R') = Tg + R$, for any coset $g + R'$.

Now consider $\theta \,\epsilon$ Hom $\{R, G\}$. Clearly the product $\theta' = \theta T$ is an element of Hom $\{R', G\}$, and the correspondence $\theta \to \theta'$ is a homomorphism T_h^* of Hom $\{R, G\}$ into Hom $\{R', G\}$. Furthermore $\theta \,\epsilon$ Hom $\{F \mid R, G\}$ implies $\theta T \,\epsilon$ Hom $\{F' \mid R', G\}$, so that T_h^* also induces a homomorphism T_h^*,

$$(12.1)\qquad T_h^* : \mathrm{Hom}\,\{R, G\}/\mathrm{Hom}\,\{F \mid R, G\} \to \mathrm{Hom}\,\{R', G\}/\mathrm{Hom}\,\{F' \mid R', G\}.$$

Similarly, consider $f \,\epsilon$ Fact $\{G, H\}$. The function f' defined by

$$f'(h', k') = f(Th', Tk') \qquad (h', k' \,\epsilon\, H')$$

is a factor set of H' in G, and the correspondence $f \to f'$ is a homomorphism T_e^* of Fact $\{G, H\}$ into Fact $\{G, H'\}$. Furthermore, $f \,\epsilon$ Trans $\{G, H\}$ implies $f' \,\epsilon$ Trans $\{G, H'\}$, so that T_e^* also induces a homomorphism T_e^* for the corresponding factor groups Ext = Fact/Trans,

$$(12.2)\qquad T_e^* : \mathrm{Ext}\,\{G, H\} \to \mathrm{Ext}\,\{G, H'\}.$$

By the (dual) homomorphisms induced on Ext or Hom by T we always mean these homomorphisms T_h^* and T_e^*.

Theorem 12.1. *Let T be a homomorphism of F' into F with $T(R') \subset R$, where $F \supset R$ and $F' \supset R'$ are free groups, while η (or η') is the homomorphism of* Hom $\{R, G\}$ *onto* Ext $\{G, F/R\}$ *established in the proof of Theorem* 10.1. *Then*

$$(12.3)\qquad \eta' T_h^* = T_e^* \eta,$$

where T_h^*, T_e^* are the appropriate homomorphisms induced by T on Hom *and* Ext, *respectively.*

Proof. The figure involved is

$$\begin{array}{ccc} \mathrm{Hom}\,\{R, G\} & \xrightarrow[\eta]{} & \mathrm{Ext}\,\{G, F/R\} \\ \downarrow T_h^* & & \downarrow T_e^* \\ \mathrm{Hom}\,\{R', G\} & \xrightarrow[\eta']{} & \mathrm{Ext}\,\{G, F'/R'\} \end{array}$$

The correspondence η was constructed from a factor set f_0 for F as an extension of R; similarly, η' is based on a factor set f_0' for H' in R', such that

$$u_0'(h') + u_0'(k') = u_0'(h' + k') + f_0'(h', k'), \tag{12.4}$$

where $u_0'(h')$ is a representative of $h' \in H'$ in F'. First we determine the relation between f_0 and f_0'. The given homomorphism T carries F' into F, H' into H and thus $u_0'(h')$ into $Tu_0'(h')$, a representative in F of Th' in H. This representative will differ from the given representative $u_0(Th')$ by an element of R, so that

$$Tu_0'(h') = u_0(Th') + \rho(h') \qquad \text{(all } h' \text{ in } H').$$

where each $\rho(h')$ lies in R. Now the representatives $Tu_0'(h')$ will add with a factor set $Tf_0'(h', k')$, as may be seen by applying T to both sides of (12.4). This factor set in associate (in the group TH') to the originally given factor set f_0 of $H \supset TH'$; explicitly we have, by the argument leading to (7.4), that

$$Tf_0'(h', k') = f_0(Th', Tk') + [\rho(h') + \rho(k') - \rho(h' + k')].$$

Suppose now that $\theta \in \mathrm{Hom}\ \{R, G\}$ is given. Application of η and then T_e^* will give, by the definitions of these correspondences, a factor set f', with

$$\begin{aligned} f'(h', k') &= \theta[f_0(Th', Tk')] \\ &= \theta T[f_0'(h', k')] + [\theta\rho(h' + k') - \theta\rho(h') - \theta\rho(k')]. \end{aligned}$$

On the other hand, application of T_h^* and then η' will give, again by the appropriate definitions, a factor set f^* with

$$f^*(h', k') = \theta'[f_0'(h', k')] = \theta T[f_0'(h', k')].$$

Since $\theta\rho(h')$ is an element in G for each $h' \in H'$, these two equations show that f^* and f' are associate, hence that $f' = T_e^*\eta\theta$ and $f^* = \eta' T_h^* \theta$ do determine the same element of Ext $\{G, H\}$, as asserted in the theorem.

CHAPTER III. EXTENSIONS OF SPECIAL GROUPS

In this chapter we shall determine Ext $\{G, H\}$ more explicitly for various special groups G and H. We begin with a brief review of the theory of characters, which will be used extensively in this chapter and also in Chapters V and VI.

13. Characters[16]

Let G, H, and J be three generalized topological groups. G and H are said to be *paired* to J if a continuous function[17] $\phi(g, h)$ with values in J is given

[16] The character theory was discovered by Pontrjagin (see [8]), generalized by van Kampen (see Weil [12], Ch. VI and Lefschetz [7] Ch. II).

[17] As a mapping $G \times H \to J$; for discussion of pairing, cf. [8], [14].

such that for any fixed g_0, $\phi(g_0, h)$ is a homomorphism of H into J and for any fixed h_0, $\phi(g, h_0)$ is a homomorphism of G into J.

Each subset $A \subset G$ determines a corresponding subset Annih $A \subset H$, called the *annihilator* of A, such that $h \in$ Annih A if and only if $\phi(g, h) = 0$ for all $g \in A$. Annihilators of subsets of H are defined similarly. It is clear that the annihilators are subgroups.

LEMMA 13.1. *If G and H are paired to a topological group J, then for each $A \subset G$,* Annih A *is a closed subgroup of H.*

This is an immediate consequence of the continuity of ϕ for fixed g.

G and H are said to be *dually paired* to J if they are so paired that

$$\text{Annih } G = 0 \quad \text{and} \quad \text{Annih } H = 0.$$

LEMMA 13.2. *If G and H are paired to J then $G/$*Annih *H and $H/$*Annih *G are dually paired to J.*

The most important group pairings arise when $J = P$ is the additive group of reals reduced modulo 1. A homomorphism of a group G into P will be called a *character* and the group Hom $\{G, P\}$ will be written as Char G. Since P has no "arbitrarily small" subgroups, it follows from a remark in §3 that if G is compact, Char G is discrete. Vice versa, by Corollary 3.2, if G is discrete, Char G is compact and topological.

The basic lemma of the theory of characters is

LEMMA 13.3. *Let G be a discrete or compact topological group and let $g \neq 0$ be an element of G. There is then a character $\theta \in$* Char *G such that $\theta(g) \neq 0$.*

In the case of discrete G the lemma follows easily from the proof of Corollary 11.4, since P is infinitely divisible. In the compact case the proof is much less elementary and uses the theory of invariant integration in compact groups.

The lemma can be equivalently formulated as follows:

LEMMA 13.4. *Let G be a discrete or compact topological group. G and* Char *G are dually paired to P with the multiplitation*

$$\phi(g, \theta) = \theta(g), \qquad g \in G,\ \theta \in \text{Char } G.$$

Now let G and H be paired to P with $\phi(g, h)$ as multiplication. Since, for a fixed g, $\phi(g, h)$ is a character of H and, for fixed h, $\phi(g, h)$ is a character of G, we obtain induced mappings

$$G \to \text{Char } H, \qquad H \to \text{Char } G. \tag{13.1}$$

A basic result of the character theory is

THEOREM 13.5. *Let the compact topological group G and the discrete group H be paired to P. The pairing is dual if and only if the induced mappings* (13.1) *are isomorphisms:*

$$G \cong \text{Char } H \quad \textit{and} \quad H \cong \text{Char } G.$$

The following two theorems are consequences of the previous results:

THEOREM 13.6. *If G is a discrete or a compact topological group, then*

$$\text{Char Char } G \cong G.$$

THEOREM 13.7. *If the compact topological group G and the discrete group H are dually paired to P, then for every closed subgroup G_1 of G and every subgroup H_1 of H we have*

$$\text{Annih [Annih } G_1] = G_1, \qquad \text{Annih [Annih } H_1] = H_1.$$

14. Modular traces

To study Ext $\{G, H\}$ for compact G we need a certain modification of the "trace" of an endomorphism of a free group. The simplest case of this modification refers to a correspondence which is not a homomorphism, but is a homomorphism, modulo m-folds of elements. It may be stated as follows.

LEMMA 14.1. *Let m be an integer, and let $r \to S(r)$ be a correspondence carrying the free group R into a finite subset of itself in such manner that*

$$S(r_1 + r_2) \equiv S(r_1) + S(r_2) \qquad (\text{mod } mR), \tag{14.1}$$

for all $r_1, r_2 \in R$. Let the elements y_α be any independent basis for R, and write $S(y_\alpha) = \sum_\beta c_{\alpha\beta} y_\beta$, with integral coefficients $c_{\alpha\beta}$. Then the "trace"

$$t_m(S) \equiv \sum_\alpha c_{\alpha\alpha} \qquad (\text{mod } m) \tag{14.2}$$

is a well defined finite integer, modulo m, independent of the choice of the basis y_α for R.

The proof is exactly parallel to the standard one (e.g. [1], p. 569) for an actual homomorphism of R to itself, using the "modular" homomorphism condition (14.1) at the appropriate junctures in place of the full homomorphism condition. A similar analogue of a special case of the "additivity" of traces will give the following conclusion.

LEMMA 14.2. *If in Lemma 14.1 the elements $w_1, \cdots, w_t$ are any independent elements of R such that S(R) lies in the group generated by $w_1, \cdots, w_t$, and if $S(w_i) = \sum_j d_{ij} w_j$, then $t_m(S) \equiv \sum_i d_{ii}$ (mod m).*

Now let R be a subgroup of the free group F, σ a homomorphism of R into a finite subgroup of F/R. There will then be at least one integer m for which $m\sigma(R) = 0$. Choose for each coset u of F/R a representative $\rho(u)$ in F; then $\rho(u + v) \equiv \rho(u) + \rho(v)$ (mod R). For each $r \in R$, $m(\rho\sigma r)$ is also an element of R, and $S(r) = m(\rho\sigma r)$, where

$$R \xrightarrow[\sigma]{} F/R \xrightarrow[\rho]{} F \xrightarrow[m]{} R,$$

is a correspondence of R to R with the modular homomorphism property (14.1).[18] The trace of the original homomorphism σ is now defined as

$$t(\sigma) \equiv t_m(S)/m \equiv t_m(m\rho\sigma)/m \qquad (\text{mod } 1). \tag{14.3}$$

[18] S could also be described in terms of m and σ as follows: S is the essentially unique correspondence of R to a finite subset of $mF \cap R$ with property (14.1) and such that each $\sigma(r)$ is the coset modulo R of $S(r)/m$.

THEOREM 14.3. *If $R \subset F$, F a free group, and if σ is any homomorphism of R into a finite subgroup of F/R, then the trace $t(\sigma)$ defined by (14.3) is a unique real number, modulo 1, independent of the choices of m and ρ made in its definition. If σ_1 and σ_2 are two such homomorphisms of R to F/R,*

$$t(\sigma_1 + \sigma_2) \equiv t(\sigma_1) + t(\sigma_2) \quad (mod\ 1). \tag{14.4}$$

In particular, $t(_0) \equiv 0$ (mod 1). Furthermore, if T_0 is a fixed finite subgroup of F/R, the correspondence $\sigma \to t(\sigma)$ is a continuous homomorphism of Hom $\{R, T_0\}$ *into the reals modulo* 1.

We are to prove the invariance of the definition of t. First, hold ρ fixed and replace m by a proper multiple $m' = km$. Then S and $t_m(S)$ are both multiplied by k, hence $t'(\sigma) \equiv t_{km}(kS)/km \equiv kt_m(S)/km \equiv t(\sigma)$ is unaltered, mod 1. Now hold m fixed and let ρ' be any second set of representatives $\rho'(u)$ for cosets $u \in F/R$. Then $\rho'(u) \equiv \rho(u)$ (mod R), so $S'(r) \equiv S(r)$ (mod mR), which implies that $t_m(S') \equiv t_m(S)$ (mod m). This shows that the trace is independent of ρ and m.

The additive property (14.4) is readily established; it is only necessary to choose a single integer in such a way that both $m\sigma_1 R$ and $m\sigma_2 R$ are zero.

Before establishing the continuity of $t(\sigma)$, we propose a more explicit representation of the finiteness of $t(\sigma)$. Let T_0 be a fixed finite subgroup of F/R, and choose a direct summand F_0 of F with a finite number of generators such that $F_0/(F_0 \cap R)$ contains T_0. We can choose simultaneously ([1], p. 566) a basis $z_1, \cdots, z_n$ for F_0 and a basis $y_1, \cdots, y_s$ for $F_0 \cap R$ so that $y_i = d_i z_i$, for integers d_i, $i = 1, \cdots, s \leqq n$. Furthermore, one can prove $F_0 \cap R$ a direct summand of R; there is then a (not necessarily denumberable) basis for R of the form $y_1, \cdots, y_s, y_\alpha, y_\beta, \cdots$. In particular, if $\sigma(R) \subset T_0$, we may choose $\rho(0) = 0$, $\rho(T_0) \subset F_0$, hence $S(R) = m\rho\sigma(R) \subset F_0 \cap R$. The equations for S and its trace then take the form

$$S(y_\gamma) = \sum_{i=1}^{s} c_{\gamma i} y_i, \qquad t_m(S) \equiv \sum_{i=1}^{s} c_{ii} \quad (\text{mod } m), \tag{14.5}$$

where $\gamma = 1, 2, \cdots, s, \alpha, \beta, \cdots$.

To prove $t(\sigma)$ continuous it suffices to establish the continuity at $\sigma = 0$, and hence to prove that $t(\sigma) \equiv 0$ for σ in a suitable neighborhood U of 0 in Hom $\{R, T_0\}$. Let U be the open set in Hom $\{R, T_0\}$ consisting of all σ with $\sigma(y_1) = \cdots = \sigma(y_s) = 0$, where y_i is the special basis constructed from F_0 above. Then, because $\rho(0) = 0$, we have $S(y_i) = 0$, $t_m(S) \equiv 0$ (mod m), and therefore $t(\sigma) \equiv 0$ (mod 1) for σ in U.

We next consider circumstances under which the traces will vanish.

LEMMA 14.4. *If $\sigma \in$ Hom $\{R, F/R\}$ has an extension σ^* which carries F homomorphically into a finite subgroup T_0 of F/R, then $t(\sigma) \equiv 0$ (mod 1).*

PROOF. For T_0 we choose $y_i = d_i z_i$ as above, and then select ρ with $\rho(T_0) \subset F_0$ and m with $mT_0 = 0$ and each $d_i \equiv 0$ (mod m). Then, for suitable integers e_{ij},

$$\rho\sigma^*(z_i) = \sum_{j=1}^{s} e_{ij} z_j, \qquad i = 1, \cdots, n;$$

furthermore $\rho\sigma^*(kz_i) \equiv k\rho\sigma^*(z_i) \pmod{R_0}$, for any integer k. But $S(y_i) = m\rho\sigma(y_i) = m\rho\sigma^*(d_iz_i) \equiv md_i\rho\sigma^*(z_i) \pmod{mR_0}$. Then computing $t_m(S)$ by (14.5) and using the fact that $m \equiv 0 \pmod{d_j}$ for each j, we find that $t_m(S) \equiv m \sum e_{ii} \equiv 0 \pmod{m}$, as asserted.

Conversely, we can find certain circumstances in which the trace will assuredly not vanish.

LEMMA 14.5. *If $z \in F$ has order n, modulo R, and if σ is a homomorphism of R into the subgroup of F/R generated by the coset of z, then $\sigma(nz) \neq 0$ implies $t(\sigma) \not\equiv 0$ (mod 1).*

PROOF. Let u denote the coset of z, modulo R. Choose the system of representatives so that $\rho(iu) = iz$, for $i = 0, \cdots, n-1$, and use n as the integer m in the definition of the trace. Then $S = m\rho\sigma$ carries R into the cyclic subgroup generated by mz. Since $\sigma(nz) = ku$, where $k \not\equiv 0 \pmod{m}$, $S(nz) \equiv knz$, and the trace, as computed by Lemma 14.2, is $t_m(S) \equiv k \not\equiv 0 \pmod{m}$, as asserted.

15. Extensions of compact groups

The group of extensions of a compact topological group G can be expressed as an appropriate character group.

THEOREM 15.1. *If G is compact and topological, H discrete, then* $\mathrm{Ext}_f\,\{G, H\} = 0$ *and there is a (bicontinuous) isomorphism*:

$$\text{(15.1)} \qquad \mathrm{Ext}\,\{G, H\} \cong \mathrm{Char\ Hom}\,\{G, H\}.$$

If G_0 is the component of 0 in G and T the subgroup of all elements of finite order in H, then also

$$\mathrm{Ext}\,\{G, H\} \cong \mathrm{Char\ Hom}\,\{G, T\} \cong \mathrm{Char\ Hom}\,\{G/G_0, T\}.$$

The last conclusion follows at once from the first, for Hom $\{G, H\}$ includes only continuous homomorphisms ϕ of the compact group G; every such homomorphism must map the connected subgroup G_0 into 0. Furthermore each ϕ carries G into a finite subgroup of the discrete group H, hence into a subgroup of T. Observe also that H is discrete, hence has no arbitrarily small subgroups; therefore (cf. §3) Hom $\{G, H\}$ is discrete, as should be the case for a character group of the compact group Ext $\{G, H\}$.

It remains to prove (15.1). Represent H as F/R; then, according to the fundamental theorem of Chapter II, (15.1) is equivalent to

$$\text{(15.2)} \qquad \mathrm{Hom}\,\{R, G\}/\mathrm{Hom}\,\{F \mid R, G\} \cong \mathrm{Char\ Hom}\,\{G, F/R\}.$$

According to Theorem 13.5 it will thus suffice to provide a suitable pairing of the compact group Hom $\{R, G\}$ and the discrete group Hom $\{G, F/R\}$ to the reals modulo 1. To this end, take any $\theta \in$ Hom $\{R, G\}$ and $\phi \in$ Hom $\{G, F/R\}$. As just above, $\phi(G)$ is a finite subgroup of F/R. Therefore $\sigma = \phi\theta$ is a homomorphism of R into a finite subgroup of F/R, so that the trace introduced in the previous section can be used to define

(15.3) $$t(\theta, \phi) \equiv t(\phi\theta) \pmod{1}.$$

We propose to show that this is the requisite pairing.

In the first place, this product is additive, for

$$t(\theta + \theta', \phi) \equiv t(\theta, \phi) + t(\theta', \phi) \pmod{1},$$

$$t(\theta, \phi + \phi') \equiv t(\theta, \phi) + t(\theta, \phi') \pmod{1}$$

follow from the corresponding property (14.4) for $\sigma = \phi\theta$. Secondly, if ϕ is fixed, the product $t(\theta, \phi)$ is continuous in θ. For when ϕ is fixed, $\sigma = \phi\theta$ maps R into a fixed finite subgroup of F/R. Since $\theta \to \phi\theta = \sigma$ is continuous, and since $\sigma \to t(\sigma)$ is continuous, by Theorem 14.3, the continuity of $t(\theta, \phi)$ follows.

As to the annihilators under this pairing, we assert that

(15.4) $$\text{Annih Hom } \{G, F/R\} = \text{Hom } \{F \mid R, G\}.$$

For suppose first that $\theta \in \text{Hom } \{F \mid R, G\}$ and let θ^* be an extension of θ to F. Then $\sigma^* = \phi\theta^*$ is an extension of $\sigma = \phi\theta$ to F, and σ^* still carries F into (the same) finite subgroup of F/R. Therefore, by Lemma 14.4, $t(\theta, \phi) \equiv t(\sigma) \equiv 0 \pmod{1}$. Hence θ is in the annihilator in question.

Conversely, let θ be fixed, and suppose that $t(\theta, \phi) \equiv 0 \pmod{1}$ for every ϕ; then $\theta \in \text{Hom } \{F \mid R, G\}$. Since G is compact and topological, it will suffice by Lemma 5.4 to prove that $\theta \in \text{Hom}_f \{R, G; F\}$. If this were not the case, there would be in F an element z of some order n, modulo R, such that $\theta(nz) = g_0$ is not an element of nG. But nG is a continuous image (under $g \to ng$) of the compact group G, hence (Lemma 1.1) is a closed subgroup of G; therefore G/nG is compact and topological. By Lemma 13.3 there is then character χ of G/nG with $\chi(g_0') \neq 0$, where g_0' is the coset of g_0 modulo nG. Since every coset of G/nG has as order some divisor of n, this character χ carries G/nG into the group generated by the fraction $1/n$, modulo 1. This is a cyclic group of order n, and so can be replaced by the isomorphic cyclic group of order n generated by the coset z' of z in F/R. The so-modified character X of G/nG then induces a continuous homomorphism ϕ of G into F/R, where

$$\phi(g_0) \neq 0, \qquad \phi(G) \subset [0, z', z'^2, \cdots, z'^{n-1}].$$

For this particular ϕ, the homomorphism $\sigma = \phi\theta$ carries nz into $\phi\theta(nz) = \phi(g_0) \neq 0$. Lemma 14.5 of the previous section then shows that $t(\sigma) \equiv t(\theta, \phi) \not\equiv 0 \pmod{1}$, contrary to the assumption $t(\theta, \phi) \equiv 0$ for every ϕ. Therefore θ does lie in Hom $\{F \mid R, G\}$, and 15.4 is proved.

Finally, we assert that, under the pairing t,

(15.5) $$\text{Annih Hom } \{R, G\} = 0.$$

For suppose instead that $t(\theta, \phi) \equiv 0 \pmod{1}$ for all θ and for some $\phi \neq 0$. Then for some $g_0 \in G$, $\phi(g_0) = u \neq 0$. The element u of F/R is the coset of some element w of F; as before, ϕ maps G into a finite subgroup of F/R, so that w

has a finite order m, modulo R. It is then possible to select in the free group F an independent basis with a first element z_0 such that $w = kz_0$ for some integer k. If z_0 has order n, modulo R, there is then a corresponding basis for R of elements y_α, with $y_0 = nz_0$. Now construct $\theta \in \text{Hom}\,\{R, G\}$ by setting

$$\theta(y_0) = g_0, \qquad \theta(y_\alpha) = 0, \qquad y_\alpha \neq y_0.$$

This particular homomorphism carries R into the subgroup of G generated by g_0, so that the product $\sigma = \phi\theta$ carries R into the finite subgroup of F/R generated by $\phi(g_0) = u$. Since u is the coset of $w = kz_0$, this is contained in the subgroup of F/R generated by the coset of z_0. Furthermore $\sigma(nz_0) = u \neq 0$. Lemma 14.5 again applies to show that $t(\sigma) \equiv t(\theta, \phi) \not\equiv 0 \pmod 1$, counter to assumption.

Given the assertions (15.4) and (15.5) as to annihilators, it follows from Lemma 13.2 that the groups $\text{Hom}\,\{R, G\}/\text{Hom}\,\{F \mid R, G\}$ and $\text{Hom}\,\{G, F/R\}$ are dually paired. Formula (15.2) is then a consequence of Theorem 13.5.

16. Two lemmas on homomorphisms

A generalized topological group G is said to have no arbitrarily small subgroups if there is in G an open set V containing 0 but containing no subgroups other than the group consisting of 0 alone.

LEMMA 16.1. *If the discrete group T has no elements of infinite order and the generalized topological group G has no arbitrarily small subgroups, while G_0 is the same group with the discrete topology, then* $\text{Hom}\{T, G\}$ *and* $\text{Hom}\{T, G_0\}$ *have the same topology.*

PROOF. $\text{Hom}\,\{T, G\}$ and $\text{Hom}\,\{T, G_0\}$ are algebraically identical. The hypotheses on T insure that every finite set of elements of T generates a finite subgroup of T. A complete set of neighborhoods U of 0 in $\text{Hom}\,\{T, G\}$ may therefore be found thus: take a finite subgroup $T_0 \subset T$ and an open set V_0 in G containing 0, and let U consist of all homomorphisms θ with $\theta(T_0) \subset V_0$. In particular, if V_0 is contained in the special open set V of G which contains no proper subgroups, the subgroup $\theta(T_0)$ is zero, so that U consists of all θ with $\theta(T_0) = 0$. The special sets U so described also form a complete set of neighborhoods of 0 in $\text{Hom}\,\{T, G_0\}$. Therefore the two topologies on the group are equivalent.

LEMMA 16.2. *Let $F \supset R$ be a free (discrete) group, $G' \supset G$ a discrete group, while* $\text{Hom}\,\{F, G'; R, G\}$ *denotes the set of all homomorphisms $\phi \in$* $\text{Hom}\,\{F, G'\}$ *with $\phi(R) \subset G$. Then*

$$\text{Hom}\,\{F, G'; R, G\}/\text{Hom}\,\{F, G\} \cong \text{Hom}\,\{F/R, G'/G\}. \tag{16.1}$$

PROOF. Any homomorphism of F/R into G'/G may be regarded as a homomorphism of F into G'/G which carries R into zero (Lemma 3.3), so that (16.1) becomes

$$\text{Hom}\,\{F, G'; R, G\}/\text{Hom}\,\{F, G\} \cong \text{Hom}\,\{F, G'/G; R, 0\}. \tag{16.2}$$

For each $\phi \in \text{Hom}\ \{F, G'\}$ let ϕ^* be the corresponding homomorphism reduced modulo G, so that for $x \in F$, $\phi^*(x)$ is the coset of $\phi(x)$, modulo G. The correspondence $\phi \to \phi^*$ is a homomorphism mapping Hom $\{F, G'; R, G\}$ into Hom $\{F, G'/G; R, 0\}$. Furthermore $\phi^* = 0$ if and only if $\phi(F) \subset G$, or $\phi \in \text{Hom}\ \{F, G\}$. Therefore $\phi \to \phi^*$ provides an (algebraic) isomorphism of the left hand group in (16.2) to a subgroup of the right hand group.

Conversely, select a fixed basis z_α for the free group F, and for each coset $b \in G'/G$ pick a fixed representative element $\rho(b)$ in G'. For given $\sigma \in \text{Hom}\ \{F, G'/G; R, 0\}$, define a corresponding homomorphism $\phi = \phi(\sigma)$, for any $x = \sum k_\alpha z_\alpha \in F$, as

$$\phi(\sum_\alpha k_\alpha z_\alpha) = \sum_\alpha k_\alpha \rho[\sigma z_\alpha].$$

This is a homomorphism of F into G'. By construction, $\rho[\sigma z_\alpha]$ modulo G is just σz_α, hence $\phi(x)$, modulo G, is $\sigma(x)$, or $\phi^* = \sigma$. This implies that $\phi(R) \subset G$, and so that each σ is the correspondent of some ϕ in the homomorphism $\phi \to \phi^*$.

To show (16.2) bicontinuous, we first analyze the topology in the groups involved. By the definition of the topology in a factor group, we have to consider only open sets in Hom $\{F, G'; R, G\}$ which are unions of cosets of Hom $\{F, G\}$. If $z_1, \cdots, z_n$ is any finite selection from the fixed set of generators for F, the set $U(z_1, \cdots, z_n)$ consisting of all ϕ with $\phi(z_1) \equiv \cdots \equiv \phi(z_n) \equiv 0 \pmod G$ is such an open set, and contains $\phi_0 = 0$. We assert that any open set V containing 0 which is a union of cosets contains one of these sets U. For, given V, there will be elements $x_1, \cdots, x_m$ in F such that V contains all ϕ with $\phi x_i = 0$. Select generators $z_1, \cdots, z_n$ such that each x_i can be expressed in terms of $z_1, \cdots, z_n$; then V contains all ϕ with $\phi z_i = 0$. Moreover, if $\phi z_i \equiv 0 \pmod G$, there is a homomorphism ϕ_1 of F into G with $\phi z_i = \phi_1 z_i$; since $\phi - \phi_1 \in V$, since V is a union of cosets of Hom $\{F, G\}$, and since $\phi_1 \in \text{Hom}\ \{F, G\}$, we conclude that $\phi \in V$. Thus $V \supset U(z_1, \cdots, z_n)$.

A similar but simpler argument for Hom $\{F, G'/G; R, 0\}$ will show that every open set containing zero in this group contains all σ with $\sigma z_1 = \cdots = \sigma z_n = 0$, for a suitable set of the generators of F. The mapping $\sigma \to \phi$ carries open sets of this special type into the open sets $U(z_1, \cdots, z_n)$ described above, and conversely. This shows that the correspondence $\phi \to \phi^*$ is continuous at 0, and hence everywhere.

17. Extensions of integers

Next we consider the case in which every element of H has finite order; we then write T instead of H for this group. The group of extensions of the integers by such a group T can be written as a group of characters. In case T is finite, the result is a generalization of Corollary 11.2, for in this case Char $T \cong T$.

THEOREM 17.1. *If T has only elements of finite order, and if I is the (additive) group of integers,*

$$\text{Ext}_f\ \{I, T\} = 0, \tag{17.1}$$

(17.2) $$\text{Ext}\ \{I, T\} \cong \text{Char}\ T.$$

The methods used to establish this result apply with equal force if I is replaced by any discrete group G which has no elements of finite order. The group Char T of homomorphisms of T into the group of reals modulo 1 must then be replaced by a group of homomorphisms of T into another group suitably constructed from G. In fact, any G with no elements of finite order can be embedded in an essentially unique discrete group G_∞ with the following properties:[19]

(i) G_∞ has no elements of finite order,

(ii) G_∞/G has only elements of finite order,

(iii) G_∞ is infinitely divisible.

For any $g \in G_\infty$ and any integer m there is then a unique $h = g/m$ in G_0 with $mh = g$. The (discrete) factor group G_∞/G is the analogue of the topological group P' of rationals modulo 1. Specifically, if $G = I$, $G_\infty = I_\infty$ is the group of rational numbers, and G_∞/G is the group P', but with a discrete topology. Since T has only elements of finite order, Char T is Hom $\{T, P'\}$. But P' clearly has no arbitrarily small subgroups, so that the latter group, by Lemma 16.1, is identical (algebraically and topologically) with Hom $\{T, I_\infty/I\}$. The exact generalization of Theorem 17.1 is thus

THEOREM 17.2. *If T has only elements of finite order, while G is discrete and has no elements of finite order, and G_∞ is defined as above,*

(17.3) $$\text{Ext}_f\ \{G, T\} = 0,$$

(17.4) $$\text{Ext}\ \{G, T\} \cong \text{Hom}\ \{T, G_\infty/G\}.$$

The isomorphism is bicontinuous if G and G_∞/G are both discrete.

PROOF. If T is represented in the form $T = F/R$, for F free, the conclusions of this theorem can be reformulated, according to the fundamental theorem of Chapter II, as

(17.3a) $$\text{Hom}_f\ \{R, G; F\} = \text{Hom}\ \{F \mid R, G\},$$

(17.4a) $$\text{Hom}\ \{R, G\}/\text{Hom}\ \{F \mid R, G\} \cong \text{Hom}\ \{F/R, G_\infty/G\}.$$

Observe first that any homomorphism $\theta \in \text{Hom}\ \{R, G\}$ can be extended in a unique way to a homomorphism $\theta^* \in \text{Hom}\ \{F, G_\infty\}$. For, since every element of $T = F/R$ has finite order, every $z \in F$ has a finite order modulo R. For each such z pick an integer m such that $mz \in R$, and define

(17.5) $$\theta^*(z) = (1/m)\theta(mz), \qquad z \in F,\ mz \in R.$$

This definition of θ^* is independent of the choice of m, and does yield a homomorphism of F into G_∞. Clearly it is the only such homomorphism extending the given θ.

Suppose now that $\theta \in \text{Hom}_f\ \{R, G; F\}$. Each element $z \in F$ then generates a

---

[19] G_∞ could also be described as a tensor product; see §18.

finite subgroup of F/R, so θ can be extended to a homomorphism mapping z and R into G. This extension of θ must agree with the unique extension θ^*. This shows that $\theta^*(z) \in G$ for each z, so that θ^* is in fact a homomorphism of F into $G \subset G_\infty$, and $\theta \in \text{Hom}\,\{F \mid R, G\}$. This proves (17.3a).

As in §16, let Hom $\{F, G_\infty ; R, G\}$ denote the group of all homomorphisms $\phi \in \text{Hom}\,\{F, G_\infty\}$ with $\phi(R) \subset G$. This is a topological group, under the usual specification (§1) that any open set in Hom $\{F, G_\infty ; R, G\}$ is the intersection of this group with an open set in the topological group Hom $\{F, G_\infty\}$.

The correspondence $\phi \to \phi \mid R$ provides a bicontinuous isomorphism

$$\text{(17.6)} \qquad \text{Hom}\,\{F, G_\infty ; R, G\} \cong \text{Hom}\,\{R, G\}.$$

For, by Lemma 3.4, $\phi \to \phi \mid R$ is a continuous homomorphism. It is an isomorphism because each $\theta \in \text{Hom}\,\{R, G\}$ has a unique extension $\theta^* = \phi \in \text{Hom}\,\{F, G_\infty ; R, G\}$, by (17.5). This inverse correspondence is also continuous; for if U is the open set consisting of all ϕ with $\phi z_i = g_i$, for given $z_i \in F$ and $g_i \in G_\infty$, $i = 1, \cdots, n$, there is an open set U_m in Hom $\{R, G\}$ consisting of all θ with $\theta(mz_i) = mg_i$, where m is chosen so that each $mz_i \in R$ and each $mg_i \in G$. The correspondence $\theta \to \theta^*$ of (17.5) carries U_m into U. This proves (17.6).

The correspondence $\phi \to \phi \mid R$ maps the subgroup Hom $\{F, G\}$ of Hom $\{F, G_\infty ; R, G\}$ onto Hom $\{F \mid R, G\}$. Hence (17.6) also yields an isomorphism

$$\text{Hom}\,\{F, G_\infty ; R, G\}/\text{Hom}\,\{F, G\} \cong \text{Hom}\,\{R, G\}/\text{Hom}\,\{F \mid R, G\}.$$

On the other hand, Lemma 16.2 provides an isomorphism

$$\text{Hom}\,\{F, G_\infty ; R, G\}/\text{Hom}\,\{F, G\} \cong \text{Hom}\,\{F/R, G_\infty/G\}.$$

These two combine to give the required isomorphism (17.4a).

It should be remarked that the results of this section can also be obtained by arguments directly on factor sets, without the interposition of the fundamental theorem of Chapter II. Specifically, to prove Theorem 17.2, one could consider an extension E of G by T, determined by a factor set $f(s, t)$ for $s, t \in T$. If $t \in T$ has order m, let $\phi_E(t) \equiv (1/m)\sum_i f(it, t) \pmod{G}$, where $i = 0, 1, \cdots, m - 1$. In this fashion E determines a homomorphism $\phi_E \in \text{Hom}\,\{T, G_\infty/G\}$. Conversely, given such a homomorphism ϕ, one may select for each $\phi(t) \in G_\infty/G$ a representative element $\phi'(t) \in G_\infty$ and construct the corresponding factor set as $f(s, t) = \phi'(s) + \phi'(t) - \phi'(s + t)$. These correspondences will establish (17.4). The device of constructing ϕ_E by summation over the terms of the factor set is an application of the so-called "Japanese homomorphism," as commonly used for (multiplicative) factor sets.

18. Tensor products

Some of our formulas can be expressed more easily by means of the tensor products introduced by Whitney [13]. If A and B are given discrete abelian groups the *tensor product* $A \circ B$ is a set whose elements are finite formal sums

$\sum a_ib_i$ of formal products a_ib_i, with each $a_i \in A$, $b_i \in B$. Two such elements are added simply by combining the two formal sums into a single sum. Two such elements are equal if and only if the second can be obtained from the first by a finite number of replacements of the forms $(a + a')b \leftrightarrow ab + a'b$, $a(b + b') \leftrightarrow ab + ab'$. The tensor product $A \circ B$ so defined is a discrete abelian group, and the multiplication $a \cdot b$ is a pairing of A and B to $A \circ B$.

In the special case when $B = G$ is a group containing no elements of finite order, and $A = R_0$ is the additive group of rational numbers, any sum $\sum a_ib_i$ can, by the distributive law, be rewritten as a single term $(r/s)b$, where s is a common denominator for the rational numbers a_i. This representation is essentially unique. Therefore $R_0 \circ G$ is simply the group G_∞ used in §17 above, and G_∞/G is $(R_0 \circ G)/G$ (for details, cf. Whitney [13], pp. 507–508).

The tensor product can equivalently be defined in terms of characters, in the following fashion:

THEOREM 18.1. *If A and B are (discrete) abelian groups,*

$$A \circ B \cong \text{Char Hom }\{B, \text{Char } A\}. \tag{18.1}$$

PROOF. This conclusion can also be written in the form

$$\text{Char }(A \circ B) \cong \text{Hom }\{B, \text{Char } A\}. \tag{18.2}$$

Since the group of characters is the group of homomorphisms into the group P of reals modulo 1, this conclusion is a special case (with $C = P$) of the following

LEMMA 18.2. *If A and B are discrete abelian groups, C any generalized (topological) abelian group, then there is a bicontinuous isomorphism*

$$\text{Hom }\{A \circ B, C\} \cong \text{Hom }\{B, \text{Hom }(A, C)\}. \tag{18.3}$$

PROOF. Let $\theta \in \text{Hom }\{A \circ B, C\}$ be given. For each $b \in B$, let $\phi_b(a) = \theta(ab)$. Then $\phi_b \in \text{Hom }(A, C)$. Let $\omega_\theta(b) = \phi_b$. Then $\omega_\theta \in \text{Hom }\{B, \text{Hom }(A, C)\}$, and the correspondence $\theta \to \omega_\theta$ is a homomorphism of $\text{Hom }\{A \circ B, C\}$ into $\text{Hom }\{B, \text{Hom }(A, C)\}$. One verifies readily that it is an (algebraic) isomorphism ($w_\theta = 0$ only if $\theta = 0$). Furthermore, it is an isomorphism onto the whole group $\text{Hom }\{B, \text{Hom }(A, C)\}$. For let any ω in the latter group be given, with $\omega(b) = \phi_b' \in \text{Hom }(A, C)$ for each $b \in B$. Then define

$$\theta_\omega(\textstyle\sum a_ib_i) = \sum \phi_{b_i}'(a_i), \qquad a_i \in A,\ b_i \in B.$$

One verifies that θ_ω is uniquely defined, under the identifications $(a + a')b \to ab + a'b$, $a(b + b') \to ab + ab'$ used in the definition of $A \circ B$. Furthermore, $\theta_\omega \in \text{Hom }\{A \circ B, C\}$, and $\theta_\omega \to \omega$ in the previously given correspondence. Therefore $\theta \to \omega_\theta$, $\omega \to \theta_\omega$ does yield the indicated isomorphism (18.3). The continuity of the isomorphism in both directions is readily established from these explicit formulas and the appropriate definitions of open sets in the given topologies of the groups concerned.

Chapter IV. Direct and Inverse Systems

The Čech homology groups for a space are defined as limits of certain "direct" and "inverse" systems of homology groups for finite coverings of the space (Chap. VI). In view of our representation of homology groups in terms of groups of homomorphisms and groups of group extensions we are led to consider limits of groups of this sort. We shall show that the limit of a group of homomorphisms is itself a group of homomorphisms (§21) and that the corresponding proposition holds in certain special cases for groups of group extensions (§22). In the general case, however, we must introduce a new group to represent the limit of a group of group extensions. This group can also be introduced as a limit of tensor products (§25).

19. Direct systems of groups

A directed set J is a partially ordered set of elements $\alpha, \beta, \gamma, \cdots$ such that for any two elements α and β there always exists an element γ with $\alpha < \gamma$, $\beta < \gamma$. For each index α in a directed set J let H_α be a (discrete) group, and for each pair $\alpha < \beta$, let $\phi_{\beta\alpha}$ be a homomorphism of H_α into H_β. If $\phi_{\gamma\alpha} = \phi_{\gamma\beta}\phi_{\beta\alpha}$ whenever $\alpha < \beta < \gamma$, the groups H_α are said to form a *direct system* with the *projections* $\phi_{\beta\alpha}$.[20]

Any direct system determines a unique (discrete) limit group $H = \varinjlim H_\alpha$ as follows. Every element h_α of one of the groups H_α is regarded as an element h_α^* of the limit H, and two elements h_α^*, h_β^* are equal if and only if there is an index γ, $\alpha < \gamma$, $\beta < \gamma$, with $\phi_{\gamma\alpha}h_\alpha = \phi_{\gamma\beta}h_\beta$. Two elements h_α^* and h_β^* in H are added by finding some γ with $\alpha < \gamma$, $\beta < \gamma$; the sum is then the element $h_\gamma^* = (\phi_{\gamma\alpha}h_\alpha + \phi_{\gamma\beta}h_\beta)^*$. Under this addition and equality, the elements h_α^* form a group $H = \varinjlim H_\alpha$. Each of the given groups H_α has a homomorphism $\phi_\alpha(h_\alpha) = h_\alpha^*$ into the limit group, and $\phi_\beta\phi_{\beta\alpha} = \phi_\alpha$, for $\alpha < \beta$.

In case each given projection $\phi_{\beta\alpha}$ is an isomorphism (of H_α into H_β), the limit group can be regarded as a "union" of the given groups: each group H_α has an isomorphic replica $\phi_\alpha H_\alpha$ within H, and H is simply the union of these subgroups.

A subset J' of the set J of indices α is said to be *cofinal* in J if for each index α there is in J' an α' with $\alpha < \alpha'$. The limit $\varinjlim H_{\alpha'}$, taken over any such cofinal subset, is isomorphic to the original limit H.

20. Inverse systems of groups

For each index α in a directed set let A_α be a (generalized topological) group, and for each $\alpha < \beta$ let $\psi_{\alpha\beta}$ be a (continuous) homomorphism of A_β in A_α. If $\psi_{\alpha\beta}\psi_{\beta\gamma} = \psi_{\alpha\gamma}$ whenever $\alpha < \beta < \gamma$, the groups A_α are said to form an *inverse system* relative to the *projections* $\psi_{\alpha\beta}$. Each inverse system determines a limit group $A = \varprojlim A_\alpha$. An element of this limit group is a set $\{a_\alpha\}$ of elements $a_\alpha \in A_\alpha$ which "match" in the sense that $\psi_{\alpha\beta}a_\beta = a_\alpha$ for each $\alpha < \beta$. The sum

[20] Direct (and inverse) systems were discussed in Steenrod [9], Lefschetz [7], Chap. I and II, and in Weil [12], Ch. I.

of two such sets is $\{a_\alpha\} + \{b_\alpha\} = \{a_\alpha + b_\alpha\}$; since the ψ's are homomorphisms, this sum is again an element of the group. This limit group A is a subgroup of the direct product of the groups A_α. The topology of the direct product $\prod A_\alpha$ thus induces (§1) a topology in $\underleftarrow{\mathrm{Lim}}\, A_\alpha$; an open set in the latter group is the intersection with $\underleftarrow{\mathrm{Lim}}\, A_\alpha$ of an open set of $\prod A_\alpha$. This makes $\underleftarrow{\mathrm{Lim}}\, A_\alpha$ a generalized topological group. As before, a cofinal subset of the indices gives an isomorphic limit group.

Let each B_α be a subgroup of the corresponding group A_α of an inverse system, and assume, for $\alpha < \beta$, that $\psi_{\alpha\beta} B_\beta \subset B_\alpha$. Then the system B_α is an inverse system under the same projections $\psi_{\alpha\beta}$, and the limit $\underleftarrow{\mathrm{Lim}}\, B_\alpha$ is, in natural fashion, a subgroup of $\underleftarrow{\mathrm{Lim}}\, A_\alpha$. On the other hand, $\psi_{\alpha\beta}$ induces a homomorphism $\psi'_{\alpha\beta}$ of the (generalized topological) group A_β/B_β into A_α/B_α. Relative to these projections, the factor groups themselves form an inverse system A_α/B_α. The limit group of the latter system contains a homomorphic image of $\underleftarrow{\mathrm{Lim}}\, A_\alpha$; if each a_α in A_α determines a coset a'_α in A_α/B_α, the map $\{a_\alpha\} \to \{a'_\alpha\}$ is a homomorphism of $\underleftarrow{\mathrm{Lim}}\, A_\alpha$ into $\underleftarrow{\mathrm{Lim}}\, (A_\alpha/B_\alpha)$ in which exactly the elements of $\underleftarrow{\mathrm{Lim}}\, B_\alpha$ are mapped on zero. Thus we have

$$\underleftarrow{\mathrm{Lim}}\, A_\alpha/\underleftarrow{\mathrm{Lim}}\, B_\alpha \subset \underleftarrow{\mathrm{Lim}}\, (A_\alpha/B_\alpha). \tag{20.1}$$

For compact topological subgroups this is an isomorphism:

LEMMA 20.1. *If the A_α form an inverse system relative to the $\psi_{\alpha\beta}$, and if each B_α is a compact topological subgroup of A_α with $\psi_{\alpha\beta}B_\beta \subset B_\alpha$, then*

$$\underleftarrow{\mathrm{Lim}}\, A_\alpha/\underleftarrow{\mathrm{Lim}}\, B_\alpha \cong \underleftarrow{\mathrm{Lim}}\, (A_\alpha/B_\alpha). \tag{20.2}$$

PROOF. Consider any $c = \{c_\alpha\}$ in $\underleftarrow{\mathrm{Lim}}\, (A_\alpha/B_\alpha)$, where $\psi'_{\alpha\beta} c_\beta = c_\alpha$ for each $\alpha < \beta$. Each $c_\alpha \in A_\alpha/B_\alpha$ is a coset of the compact topological subgroup B_α, hence itself is a compact Hausdorff subspace of the space A_α. Furthermore $\psi_{\alpha\beta}$ is a continuous mapping of the set c_β into c_α, for each $\alpha < \beta$. Since $\psi_{\alpha\gamma} = \psi_{\alpha\beta}\psi_{\beta\gamma}$, the sets c_α form an inverse system of compact non-empty Hausdorff spaces. Their limit space is therefore[21] non-vacuous. This means that there is a set of elements $a_\alpha \in c_\alpha$ with $\psi_{\alpha\beta} a_\beta = a_\alpha$ for $\alpha < \beta$. The element $\{a_\alpha\}$ in the group $\underleftarrow{\mathrm{Lim}}\, A_\alpha$ is therefore an element which maps onto the given element $\{c_\alpha\}$ in the homomorphism $\{a_\alpha\} \to \{a'_\alpha\}$ used to establish (20.1). The continuity of (20.2), in both directions, follows readily.

There is also an "isomorphism" theorem for inverse systems.

LEMMA 20.2. *If the groups A_α form an inverse system relative to the projections $\psi_{\alpha\beta}$, while C_α form an inverse system (with the same set of indices) relative to projections $\phi_{\alpha\beta}$, and if σ_α are (bicontinuous) isomorphisms of A_α to C_α, for every α, such that the "naturality" condition $\sigma_\alpha\psi_{\alpha\beta} = \phi_{\alpha\beta}\sigma_\beta$ holds, then the groups $\underleftarrow{\mathrm{Lim}}\, A_\alpha$ and $\underleftarrow{\mathrm{Lim}}\, C_\alpha$ are bicontinuously isomorphic.*

---

[21] See Lefschetz [7], Theorem 39.1 or Steenrod [9], p. 666. Observe, however, that the latter proof is incomplete, because of the gap in lines 10-11 on p. 666.

21. Inverse systems of homomorphisms

Consider the group of all homomorphisms of H into G. As in Chap. II, §12, each projection $\phi_{\beta\alpha}$ of a direct system of groups H_α will induce a "dual" homomorphism $\phi^*_{\alpha\beta}$ of Hom $\{H_\beta, G\}$ into Hom $\{H_\alpha, G\}$. Furthermore $\phi^*_{\alpha\beta}\,\phi^*_{\beta\gamma} = \phi^*_{\alpha\gamma}$ for all $\alpha < \beta < \gamma$, so that the groups Hom $\{H_\alpha, G\}$ form an inverse system relative to these dual projections.

THEOREM 21.1. *If the (discrete) groups H_α form a direct system, then*

$$\text{Hom } \{\underrightarrow{\text{Lim}}\ H_\alpha , G\} \cong \underleftarrow{\text{Lim}} \text{ Hom } \{H_\alpha , G\}. \tag{21.1}$$

PROOF. Consider any element $\omega = \{\theta_\alpha\}$ in $\underleftarrow{\text{Lim}}$ Hom $\{H_\alpha, G\}$. To define a corresponding homomorphism θ_ω on $H = \underrightarrow{\text{Lim}}\ H_\alpha$, represent each element $h \in H$ as a projection $h = \phi_\alpha h_\alpha$ of some element $h_\alpha \in H_\alpha$, and set

$$\theta_\omega(h) = \theta_\omega(\phi_\alpha h_\alpha) = \theta_\alpha(h_\alpha), \qquad h = \phi_\alpha h_\alpha . \tag{21.2}$$

The "matching" requirement that $\theta_\alpha = \phi^*_{\alpha\beta}\theta_\beta$ for $\alpha < \beta$ readily shows that $\theta_\omega(h)$ has a unique value, independent of the representation $h = \phi_\alpha h_\alpha$ chosen. Furthermore, $\theta_\omega \in$ Hom $\{H, G\}$, and the correspondence $\omega \to \theta_\omega$ is an isomorphism.

Conversely, let any $\theta \in$ Hom $\{H, G\}$ be given, and define

$$\theta_\alpha(h_\alpha) = \theta(\phi_\alpha h_\alpha), \qquad h_\alpha \in H_\alpha . \tag{21.3}$$

If $\alpha < \beta$, $\phi^*_{\alpha\beta}\theta_\beta(h_\alpha) = \theta_\beta[\phi_{\beta\alpha}h_\alpha] = \theta[\phi_\beta\phi_{\beta\alpha}h_\alpha] = \theta(\phi_\alpha h_\alpha) = \theta_\alpha h_\alpha$; so $\phi^*_{\alpha\beta}\theta_\beta = \theta_\alpha$, and these θ's match. Therefore $\omega = \{\theta_\alpha\}$ is an element of the inverse limit group $\underleftarrow{\text{Lim}}$ Hom $\{H_\alpha, G\}$, and clearly θ_ω is the original homomorphism θ. The correspondence $\omega \to \theta_\omega$ therefore does establish the desired isomorphism (21.1). The continuity in both directions follows directly from the formulas (21.2) and (21.3) and the appropriate definition of neighborhoods of zero in the groups concerned.

22. Inverse systems of group extensions

Consider a direct system of discrete groups H_α. As in Chap. II, §12, each projection $\phi_{\beta\alpha}$ of H_α into H_β will induce a homomorphism $\phi^*_{\alpha\beta}$ of Ext $\{G, H_\beta\}$ into Ext $\{G, H_\alpha\}$. Furthermore $\phi^*_{\alpha\beta}\phi^*_{\beta\gamma} = \phi^*_{\alpha\gamma}$ for all $\alpha < \beta < \gamma$, so that the groups Ext $\{G, H_\alpha\}$ form an inverse system. Contrary to the situation in the previous section, the limit group $\underleftarrow{\text{Lim}}$ Ext $\{G, H_\alpha\}$ may not be isomorphic to Ext $\{G, \underrightarrow{\text{Lim}}\ H_\alpha\}$. An example to this effect will be given below. However, there are two important cases when "Lim" and "Ext" are interchangeable.

THEOREM 22.1. *If G is compact and topological, while the (discrete) groups H_α form a direct system, then*

$$\text{Ext } \{G, \underrightarrow{\text{Lim}}\ H_\alpha\} \cong \underleftarrow{\text{Lim}} \text{ Ext } \{G, H_\alpha\}. \tag{22.1}$$

This is proved by repeated applications of Lemma 20.1 to the representation

(22.2) $$\operatorname{Ext}\{G, H\} = \operatorname{Fact}\{G, H\}/\operatorname{Trans}\{G, H\},$$

where $H = \varinjlim H_\alpha$. Recall that any $f \in \operatorname{Trans}\{G, H\}$ has the form

$$f(h, k) = g(h) + g(k) - g(h + k), \qquad h, k \in H.$$

Here $g \in G^H$ is any mapping of H into G. Clearly $f = 0$ if and only if $g \in \operatorname{Hom}\{H, G\}$, so

(22.3) $$\operatorname{Trans}\{G, H\} \cong G^H/\operatorname{Hom}\{H, G\}.$$

The correspondence $g \to f$ is clearly continuous; since the isomorphism (22.3) is one-one and since the groups G^H and $\operatorname{Hom}\{H, G\}$ are compact, by Lemma 3.1, the bicontinuity of (22.3) follows. Furthermore, this isomorphism is a "natural" one relative to homomorphisms, so that the isomorphism theorem for inverse systems (Lemma 20.2) gives

$$\varprojlim \operatorname{Trans}\{G, H_\alpha\} \cong \varprojlim [G^{H_\alpha}/\operatorname{Hom}\{H_\alpha, G\}].$$

In this representation the groups G^{H_α} and $\operatorname{Hom}\{H_\alpha, G\}$ with the "dual" projections $\phi^*_{\alpha\beta}$ form inverse systems with the respective limits G^H and $\operatorname{Hom}\{H, G\}$. Furthermore each group $\operatorname{Hom}\{H_\alpha, G\}$ is compact and topological, so Lemma 20.1 gives

(22.4) $$\begin{aligned}\varprojlim \operatorname{Trans}\{G, H_\alpha\} &\cong \varprojlim G^{H_\alpha}/\varprojlim \operatorname{Hom}\{H_\alpha, G\}\\ &= G^H/\operatorname{Hom}\{H, G\} \cong \operatorname{Trans}\{G, H\}.\end{aligned}$$

On the other hand one may show exactly as in the proof of Theorem 21.1 on homomorphisms that there is a bicontinuous isomorphism

(22.5) $$\varprojlim \operatorname{Fact}\{G, H_\alpha\} \cong \operatorname{Fact}\{G, H\}.$$

Furthermore, each of the groups $\operatorname{Trans}\{G, H_\alpha\}$ is compact and topological, so that Lemma 20.1 applies again to prove

$$\varprojlim [\operatorname{Fact}/\operatorname{Trans}] \cong \varprojlim \operatorname{Fact}/\varprojlim \operatorname{Trans}.$$

This, with (22.4) and (22.5), gives the desired conclusion.[22]

THEOREM 22.2. *If G is discrete and has no elements of finite order, while T_α is a direct systems of discrete groups with only elements of finite order, then*

(22.6) $$\operatorname{Ext}\{G, \varinjlim T_\alpha\} \cong \varprojlim \operatorname{Ext}\{G, T_\alpha\}.$$

The proof appeals directly to the result found in Theorem 17.2 of Chapter III, to the effect that

(22.7) $$\operatorname{Ext}\{G, T_\alpha\} \cong \operatorname{Hom}\{T_\alpha, G_\infty/G\}.$$

The groups $\operatorname{Hom}\{T_\alpha, G_\infty/G\}$ will form an inverse system under the dual projections $\phi^*_{\alpha\beta}$; as in Theorem 21.1 we then have

$$\operatorname{Hom}\{\varinjlim T_\alpha, G_\infty/G\} \cong \varprojlim \operatorname{Hom}\{T_\alpha, G_\infty/G\}.$$

---

[22] Theorem 22.1 can also be proved by representing Ext by means of Char Hom $\{G, H\}$ as in Theorem 15.1. This argument, however, requires a tedious proof that the isomorphism established in the latter theorem is "natural," in the sense of §12.

But the group on the left is simply Ext $\{G, \underrightarrow{\text{Lim}}\, T_\alpha\}$, by another application of Theorem 17.2. The desired result should then follow by taking (inverse) limits on both sides in (22.7).

To carry out this argument, it is necessary to have the naturality condition which gives the isomorphism theorem (Lemma 20.2) for inverse systems. This naturality condition requires that the isomorphism (22.7) permute with the projections of the inverse systems. This is just a statement of the fact that the isomorphism (22.7) established in Theorem 17.2 is "natural" in the sense envisaged in §12. The proof of this naturality is straightforward, so details will be omitted.

COROLLARY 22.3. *If the discrete group G has only a finite number of generators, while T_α is a direct system of discrete groups with only elements of finite order, then*

$$\text{Ext}\ \{G, \underrightarrow{\text{Lim}}\ T_\alpha\} \cong \underleftarrow{\text{Lim}}\ \text{Ext}\ \{G, T_\alpha\}.$$

PROOF. Write G as $F \times L$ where F is free, L is finite (and thus compact). By (11.2) there is a "natural" isomorphism

$$\text{Ext}\ \{G, \underrightarrow{\text{Lim}}\ T_\alpha\} \cong \text{Ext}\ \{F, \underrightarrow{\text{Lim}}\ T_\alpha\} \times \text{Ext}\ \{L, \underrightarrow{\text{Lim}}\ T_\alpha\}.$$

The asserted result now follows by applying Theorem 22.2 to the first factor on the right, and Theorem 22.1 to the second, using Lemma 20.2.

We now show by an example that "Ext" and "Lim" do not necessarily commute. Let p be a fixed prime number, H the additive group of all rationals with denominator a power of p, and H_n the subgroup consisting of all multiples of $1/p^n$. Then $\underrightarrow{\text{Lim}}\, H_n = H$, since H is the union of the groups H_n. Furthermore H_n is a free group, so Ext $\{I, H_n\} = 0$, where I is the group of integers. On the other hand, Ext $\{I, \underrightarrow{\text{Lim}}\, H_n\} =$ Ext $\{I, H\}$ is a group computed in appendix B; it is decidedly not zero, in fact it is not even denumerable.

23. Contracted extensions

Before further consideration of the inverse limits of groups of extensions, we make a comparison of the group of extensions of a group G by a group H with the group of extensions by a subgroup H_0 of H. The identity mapping I of H_0 into H is a homomorphism, hence, as in §12, will give dual homorphisms

$$I^*: \text{Fact}\ \{G, H\} \to \text{Fact}\ \{G, H_0\}, \tag{23.1}$$

$$I^*: \text{Trans}\ \{G, H\} \to \text{Trans}\ \{G, H_0\}. \tag{23.2}$$

Specifically, I^* is the operation of "cutting off" a factor set $f \in$ Fact $\{G, H\}$ to give a factor set $f_0 = I^*f \in$ Fact $\{G, H_0\}$; $f_0(h, k)$ is defined only for $h, k \in H_0$, and always equals $f(h, k)$. Clearly I^* carries transformation sets into transformation sets, as in (23.2). Thus I^* also induces a dual homomorphism

$$I^*: \text{Ext}\ \{G, H\} \to \text{Ext}\ \{G, H_0\}. \tag{23.3}$$

This homomorphism may be visualized as follows: given E such that $G \subset E$

and $E/G = H$, there is an $E_0 \subset E$ such that $G \subset E_0$ and $E_0/G = H_0$. Then $I^*(E) = E_0$.

LEMMA 23.1. *If H_0 is a subgroup of the group H then for any group G the homomorphism I^* of* (23.3) *maps the group* Ext $\{G, H\}$ *onto* Ext $\{G, H_0\}$.

PROOF.[23] Represent H as F/R, where F is free. There is then a subgroup F_0 of F such that $R \subset F_0$ and $F_0/R = H_0$. By the fundamental theorem we have isomorphisms

$$\text{Ext }\{G, H\} \cong \text{Hom }\{R, G\}/\text{Hom }\{F \mid R, G\},$$

$$\text{Ext }\{G, H_0\} \cong \text{Hom }\{R, G\}/\text{Hom }\{F_0 \mid R, G\},$$

where Hom $\{F \mid R, G\} \subset$ Hom $\{F_0 \mid R, G\}$. According to the "naturality" theorem of §12 the homomorphism I^* between the groups on the left can be represented on the right as that correspondence which carries each coset of Hom $\{F \mid R, G\}$ into the coset of Hom $\{F_0 \mid R, G\}$ in which it is contained. This makes it obvious that the homomorphism is a mapping "onto."

LEMMA 23.2. *If $H_0 \subset H$, then the dual homomorphisms I^* of factor and transformation sets, as in* (23.1) *and* (23.2), *are mappings "onto."*

PROOF. Any element in Trans $\{G, H_0\}$ has the form

$$f(h, k) = g(h) + g(k) - g(h + k),$$

where g is an arbitrary function on H_0 to G. Let g^* be an arbitrary extension of g to H, and

$$f^*(h, k) = g^*(h) + g^*(k) - g^*(h + k).$$

Then f^* is a transformation set with $I^*f^* = f$. This proves that (23.2) is a mapping onto. Since (23.3) and (23.2) are mappings onto, the same holds for (23.1).

24. The group Ext*

Since limits do not always permute with groups of extensions, we now introduce a new group which is the limit of an inverse system of groups of group extensions.

Consider a discrete group T with only elements of finite order. The set $\{S_\alpha\}$ of all finite subgroups of T is a direct system, if $\alpha < \beta$ means that $S_\alpha \subset S_\beta$, and that the projection $I_{\beta\alpha}$ of S_α into S_β is simply the identity. The direct limit of $\{S_\alpha\}$ is the group T.

Let G be any generalized topological group. Since $\{S_\alpha\}$ is a direct system, it follows from a previous section that the groups Ext $\{G, S_\alpha\}$ form an inverse system with projections $I^*_{\alpha\beta}$. We define our new group as the limit of this system

[23] The lemma can also be proved directly in terms of the group extensions E, E_0, using a suitable transfinite induction.

(24.1) $$\mathrm{Ext}^* \{G, T\} = \varprojlim \mathrm{Ext} \{G, S_\alpha\}.$$

The two theorems of §22 as to cases in which "Ext" and "Lim" commute give at once

COROLLARY 24.1. *If G is compact and topological, or is discrete without elements of finite order, then*

$$\mathrm{Ext}^* \{G, T\} \cong \mathrm{Ext} \{G, T\}.$$

In the definition of Ext* we used the approximation of T by its finite subgroups S_α. However, any approximation by finite groups will give the same result:

THEOREM 24.2. *If T_α is any direct system of finite groups, the corresponding inverse system of groups* Ext $\{G, T_\alpha\}$ *has a limit*

(24.2) $$\varprojlim \mathrm{Ext} \{G, T_\alpha\} \cong \mathrm{Ext}^* \{G, \varinjlim T_\alpha\}.$$

PROOF. In case T_α is the system of all finite subgroups of the limit $T = \varinjlim T_\alpha$, this equation is simply the definition of Ext*. In general, $T = \varinjlim T_\alpha$ is a group in which every element has finite order. Each T_α has a homomorphic projection $T'_\alpha = \phi_\alpha T_\alpha$ into the limit T, and T is simply the union of these subgroups T'_α. The set of these subgroups T'_α is therefore cofinal in the set of all finite subgroups of T. The inverse system of the groups Ext $\{G, T'_\alpha\}$, relative to the "identity" projections $I^*_{\alpha\beta}$, is cofinal in the inverse system used to define Ext*, hence gives the same limit group,

(24.3) $$\mathrm{Ext}^* \{G, T\} \cong \varprojlim \mathrm{Ext} \{G, T'_\alpha\}.$$

An element f^* in this limit group can be represented (but not uniquely) as a set $\{f_\alpha\}$ of factor sets $f_\alpha \epsilon$ Fact $\{G, T'_\alpha\}$ which "match" modulo transformation sets. This means that for each $\beta > \alpha$ there is a transformation set $t_{\alpha\beta} \epsilon$ Trans $\{G, T'_\alpha\}$ such that

$$f_\alpha(h', k') = f_\beta(h', k') + t_{\alpha\beta}(h', k'), \qquad h', k' \epsilon T'_\alpha .$$

Now each homomorphism ϕ_α of T_α into T'_α determines, as in §12, a dual homomorphism ϕ^*_α of Fact $\{G, T'_\alpha\}$ into Fact $\{G, T_\alpha\}$, defined so that $e_\alpha = \phi^*_\alpha f_\alpha$ is the factor set given by the equations

(24.4) $$e_\alpha(h, k) = f_\alpha(\phi_\alpha h, \phi_\alpha k), \qquad h, k \epsilon T_\alpha .$$

If the f_α match, one readily proves that the corresponding e_α also match, modulo transformation sets. If the representation of f^* by $\{f_\alpha\}$ is changed by adding to each f_α a transformation set, the e_α's are changed accordingly by transformation sets. Therefore the correspondence

(24.5) $$f^* = \{f_\alpha\} \rightarrow e^* = \{\phi^*_\alpha f_\alpha\} = \omega f^*$$

carries each element f^* in $\varprojlim$ Ext $\{G, T'_\alpha\}$ into a well defined element e^* in $\varprojlim$ Ext $\{G, T_\alpha\}$. One verifies at once that this correspondence is a homomorphism.

Now we use the assumption that each T_α is finite. If $\phi_\alpha h_\alpha = 0$ for some $h_\alpha \in T_\alpha$, the definition of equality in a direct system shows that $\phi_{\beta\alpha} h_\alpha = 0$ for some $\beta > \alpha$. Since the whole group T_α is finite, we can select a single $\beta = \beta_0(\alpha) > \alpha$ which will do this for all h_α, so that

$$\phi_\alpha h_\alpha = 0 \quad \text{implies} \quad \phi_{\beta\alpha} h_\alpha = 0, \qquad \beta = \beta_0(\alpha).$$

Since $\phi_\beta \phi_{\beta\alpha} = \phi_\alpha$, ϕ_β is now an *isomorphism* of $\phi_{\beta\alpha} T_\alpha$ onto T'_α. Let ϕ_β^{-1} denote the inverse correspondence.

Next we show that ω, as defined by (24.5), is an isomorphism. For suppose $\omega f^* = 0$; every $\phi^*_\alpha f_\alpha$ is then a transformation set t_α. Using (24.4) and $\beta = \beta_0(\alpha)$, we then have, for any $h', k' \in T'_\alpha$,

$$f_\alpha(h', k') \equiv f_\beta(h', k') = e_\beta(\phi_\beta^{-1} h', \phi_\beta^{-1} k') = t_\beta(\phi_\beta^{-1} h', \phi_\beta^{-1} k').$$

This shows that f_α is a transformation set, hence that $f^* = \{f_\alpha\} = 0$ in Ext* $\{G, T\}$.

To construct a correspondence inverse to ω, let $e^* = \{e_\alpha\}$ be a given element in $\varprojlim$ Ext $\{G, T_\alpha\}$, where each $e_\alpha \in$ Fact $\{G, T_\alpha\}$. Define

$$f_\alpha(h', k') = e_\beta(\phi_\beta^{-1} h', \phi_\beta^{-1} k'), \qquad \beta = \beta_0(\alpha) \tag{24.6}$$

for each $h', k' \in T'_\alpha$. Since the e_α's are known to match, we may verify that the replacement of β by any larger index γ in this definition will only alter f_α by a transformation set. To show that f_α and f_γ match properly for $\alpha < \gamma$, one then chooses $\beta > \beta_0(\alpha)$, $\beta > \beta_0(\gamma)$ in (24.6) and uses the given matching of the e_α's (modulo transformation sets). Finally, one verifies easily that the correspondence $\{e_\alpha\} \to \{f_\alpha\}$ of (24.6) is the inverse of the given correspondence ω of (24.5). This establishes the isomorphism (24.2) required in the theorem. The continuity, in both directions, follows from the formulae (24.5) and (24.6).

THEOREM 24.3. *If every element of T has finite order, the group* Ext* $\{G, T\}$ *contains an everywhere dense subgroup isomorphic to* Ext $\{G, T\}/\text{Ext}_f$ $\{G, T\}$.

This will be established by constructing a "natural" homomorphism of Ext $\{G, T\}$ into Ext* $\{G, T\}$. To this end, let E be any extension of G by T determined by a factor set f. As in §23, f may be "cut off" to give a factor set f_α for any given finite subgroup $S_\alpha \subset T$. These factor sets match properly, so $\{f_\alpha\}$ determines a definite element in the inverse limit group Ext* $\{G, T\}$. Alteration of f by a transformation set alters each f_α by the correspondingly "cut off" transformation set, hence does not alter the element $\{f_\alpha\} = f^*$ of Ext*. Therefore $f \to \{f_\alpha\}$ is a well defined homomorphism of Ext into Ext*. In case f lies in Ext_f $\{G, T\}$, each f_α is a transformation set, by the very definition of Ext_f, so that $\{f_\alpha\} = 0$. Conversely, if each f_α is a transformation set, $f \in \text{Ext}_f$. We thus have a (bicontinuous) isomorphism of Ext/Ext_f onto a subgroup of Ext*.

To show this subgroup everywhere dense in Ext* it will suffice, whatever the topology in G, to show the following: Given an element $f^* = \{f'_\alpha\}$ in Ext* $\{G, T\}$ and a finite set J_0 of indices, there exists a factor set f in Fact $\{G, T\}$ such that

$f_\alpha - f'_\alpha$ is a transformation set for every index $\alpha \in J_0$. To prove this, choose a finite subgroup S_γ which contains all the groups S_α, for $\alpha \in J_0$. By Lemma 23.2, the given factor set f'_γ can be obtained by "cutting off" a suitable factor set f, so that $f_\gamma - f'_\gamma$ is the transformation set 0. The matching condition for the f'_α then shows that each difference $f_\alpha - f'_\alpha$ is also a transformation set, for $\alpha \in J_0$. This proves the property stated above, and with it, the theorem.

In many cases the subgroup considered in Theorem 24.3 is the whole group Ext*. It follows from previous considerations that this is the case when G is compact or when G is discrete and has no elements of finite order. Another important case is that when T is countable:

THEOREM 24.4. *If T is countable then*

$$\text{Ext}^* \{G, T\} \cong \text{Ext}\{G, T\}/\text{Ext}_f \{G, T\}. \tag{24.7}$$

PROOF. Since T is countable, the system of all finite subgroups of T used to define Ext* $\{G, T\}$ may be replaced by a cofinal sequence of finite subgroups T_n with $T_1 \subset T_2 \subset \cdots \subset T_n \subset \cdots \subset T$, with the identity projections $I_n : T_n \to T_{n+1}$. Therefore Ext* $\{G, T\} = \underleftarrow{\text{Lim}}$ Ext $\{G, T_n\}$. An element e^* of this group can then be represented as a sequence $\{f_n\}$ of factor sets $f_n \in$ Fact $\{G, T_n\}$ which match, in the sense that, for some g_n,

$$f_{n+1}(h, k) = f_n(h, k) + [g_n(h) + g_n(k) - g_n(h + k)] \tag{24.8}$$

for all $h, k \in T_n$. The transformation set shown in brackets may be extended to all of T by extending g_n to a function g_n^* on T, as in Lemma 23.2. We introduce a new function $s_n(h) = g_1^*(h) + \cdots + g_{n-1}^*(h)$, for all $h \in T$, and a new family of factor sets

$$f'_n(h, k) = f_n(h, k) - [s_n(h) + s_n(k) - s_n(h + k)],$$

for $h, k \in T_n$.[24] Since f'_n differs from f_n by a transformation set, the given element e^* of Ext* has both representations $\{f_n\}$ and $\{f'_n\}$. But (24.8) also shows that f'_{n+1}, cut off at T_n, is exactly f'_n. Therefore these factor sets match exactly, and provide a composite factor set f of T in G. This factor set f is one which corresponds to the given element e^* of Ext* in the "natural" homomorphism of Ext into Ext* as constructed in Theorem 24.3, so this homomorphism maps Ext on all of Ext*, as asserted in (24.7).

25. Relation to tensor products

The group Ext* introduced in this chapter is closely related to the tensor product. Since an early form ([5]) of our results was formulated in terms of tensor products, we shall briefly state the connection. Let G be any group, A a compact zero-dimensional group, $\{A_\alpha\}$ the family of all open and closed subgroups of A. Then the groups A/A_α and $G \circ (A/A_\alpha)$ both form inverse

[24] This construction is an exact group theoretic analog of a similar matching process for chains, as devised by Steenrod ([9], p. 692).

systems. The modified tensor product $G \bullet A$ is defined as the limit of the groups $G \circ (A/A_\alpha)$.

Now let the group T with all elements of finite order be represented in terms of a free group F as $T = F/R$. Each finite subgroup S_α then has a representation F_α/R, and the fundamental theorem of Chapter II asserts that

$$\text{Ext}\,\{G, S_\alpha\} \cong \text{Hom}\,\{R, G\}/\text{Hom}\,\{F_\alpha \mid R, G\}. \tag{25.1}$$

The groups on both sides here form inverse systems, relative to the identity as projections. Furthermore, the isomorphism of (25.1) permutes with these projections, so that the limits of the two direct systems in (25.1) are also isomorphic. In view of the definition of Ext*, this gives

$$\text{Ext}^*\,\{G, T\} \cong \underleftarrow{\text{Lim}}\,[\text{Hom}\,\{R, G\}/\text{Hom}\,\{F_\alpha \mid R, G\}]. \tag{25.2}$$

Now if I is the group of integers, any element $\sigma = \sum g_i\phi_i$ in the tensor product $G \circ \text{Hom}\,\{R, I\}$ determines in natural fashion the homomorphism $\theta \in \text{Hom}\,\{R, G\}$ with $\theta(r) = \sum \phi_i(r)g_i$. By a somewhat lengthy argument, this correspondence $\sigma \to \theta$ can be used to "factor out" the G in (25.2) to give

$$\text{Ext}^*\,\{G, T\} \cong \underleftarrow{\text{Lim}}\,G \circ [\text{Hom}\,\{R, I\}/\text{Hom}\,\{F_\alpha \mid R, I\}]. \tag{25.3}$$

The group in brackets here is Ext $\{I, F_\alpha/R\}$, by the fundamental theorem on group extensions. According to Theorem 17.1 it can be expressed as Char S_α. Therefore (25.3) is[25]

$$\text{Ext}^*\,\{G, T\} \cong \underleftarrow{\text{Lim}}\,(G \circ \text{Char}\,S_\alpha). \tag{25.4}$$

But the group Char S_α can, by the theory of characters (Lemma 13.2, Theorem 13.5), be rewritten as a factor group Char T/Annih S_α, where the subgroups of the form Annih S_α in Char T are exactly the open and closed subgroups in the zero-dimensional group Char T. Thus (25.4) may be restated in terms of the modified tensor product, as

$$\text{Ext}^*\,\{G, T\} \cong G \bullet \text{Char}\,T. \tag{25.5}$$

The use of the "modified" tensor product is therefore equivalent to the use of the group Ext*.

CHAPTER V. ABSTRACT COMPLEXES

Turning now to the topological applications, we will establish the fundamental theorem on the decomposition of the homology groups of an infinite complex in terms of the integral cohomology groups of the complex. This theorem will be obtained in several closely related forms (Theorems 32.1, 32.2 and 34.2) for three different types of homology groups. The largest (or "longest") homology group is that consisting of infinite cycles, with coefficients in G, reduced modulo

---

[25] This argument requires an application of the isomorphism theorem for inverse systems, and hence rests on the fact that the isomorphism of Theorem 17.1 is "natural" in the sense of §12.

the subgroup of actual boundaries. Since the latter subgroup is not in general closed, this homology group will be only a generalized topological group. This suggests the introduction of the shorter "weak" homology group, which consists of cycles modulo "weak" boundaries; i.e. those cycles which can be regarded as boundaries in any finite portion of the complex. The fundamental theorem for this type of homology group uses the group Ext_f which has been already analyzed. Finally, the group of cycles modulo the closure of the group of boundaries gives (following Lefschetz) a homology group which is always topological; for this we derive a corresponding form of the fundamental theorem. Furthermore, the standard duality between homology and cohomology groups enables us to deduce a corresponding theorem (Theorem 33.1) for the cohomology groups with coefficients in an arbitrary discrete group G.

The fundamental theorem expresses a homology group by means of a group of homomorphisms and a group of group extensions; the latter can also be represented by groups of homomorphisms, as in the basic theorem of Chapter II. The requisite connection between cycles of the homology group and homomorphisms is provided by the Kronecker index (§29).

26. Complexes

The complexes considered here will be abstract cell complexes[26] satisfying a star finiteness condition. More precisely, we consider a collection K of abstract elements σ^q called *cells*. With each cell there is associated an integer q called the dimension of σ^q. (There is no restriction requiring the dimension to be non-negative.) To any two cells σ_i^{q+1}, σ_j^q there corresponds an integer $[\sigma_i^{q+1}:\sigma_j^q]$, called the *incidence number*. K will be called a *star finite complex* provided the incidence numbers satisfy the following two conditions:

(26.1) *Given* σ_j^q, $[\sigma_i^{q+1}:\sigma_j^q] \neq 0$ *only for a finite number of indices* i;

(26.2) *Given* σ_j^{q+1} *and* σ_k^{q-1}, $\sum_i [\sigma_j^{q+1}:\sigma_i^q][\sigma_i^q:\sigma_k^{q-1}] = 0$.

Condition (26.1) is the star finiteness condition. It insures that the summation in (26.2) is finite.

If we consider the "incidence" matrices of integers

$$A^q = \| [\sigma_j^{q+1}:\sigma_i^q] \|$$

we can rewrite the two conditions as follows:

(26.1′) $\qquad A^q$ *is column finite*;

(26.2′) $\qquad A^qA^{q-1} = 0.$

Actually we could have defined a complex as a collection of matrices $\{A^q\}$, $q = 0, \pm 1, \pm 2, \cdots$, such that (26.1′) and (26.2′) hold; we must assume then that the columns of A^q have the same set of labels as do the rows of A^{q-1}, in

---

[26] Essentially like those introduced by A. W. Tucker, for the case of finite complexes. Homology and cohomology are treated as in Whitney [14].

order to form the product $A^q A^{q-1}$. A q-cell will be then either a column of A^q or the corresponding row of A^{q-1}.

A subset L of the cells of K is called an *open subcomplex* if L contains with each q-cell all incident $(q + 1)$-cells; that is, if σ_i^q in L and $[\sigma_j^{q+1}:\sigma_i^q] \neq 0$ imply $\sigma_j^{q+1} \epsilon L$. The incidence matrix A_L^q of L is then the submatrix obtained from A^q by deleting all rows and all columns belonging to cells not in L. Conditions (26.1) and (26.2) automatically hold in L, the latter because of the requirement that L be "open."

A subset $L \subset K$ *is a closed subcomplex* if L contains with each q-cell all incident $(q - 1)$-cells; that is, if $\sigma_j^q \epsilon L$ and $[\sigma_j^q:\sigma_k^{q-1}] \neq 0$ imply $\sigma_k^{q-1} \epsilon L$. The incidence matrix of L is obtained as before, and the conditions (26.1) and (26.2) again hold in L. Whenever L is a closed subcomplex, its complement $K - L$ is an open one, and vice versa.

A subset L of K will be called *q-finite* if L contains only a finite number of q-cells. Because K is star-finite, every $(q - 1)$-cell is contained in a q-finite open subcomplex of K.

27. Homology and cohomology groups

Let G be an abelian group. A q-dimensional *chain* c^q in K with coefficients in G is a function which associates to every q-cell σ_i^q in K an element g_i of G. We write c^q as a formal infinite sum

$$c^q = \sum_i g_i \sigma_i^q.$$

The sum of two chains $\sum g_i\sigma_i^q$ and $\sum h_i\sigma_i^q$ is the chain $\sum (g_i + h_i)\sigma_i^q$, and the chains form a group denoted by $C^q(K, G)$. If $g_i \neq 0$ for only a finite number of indices i then the chain c^q is *finite*. The finite chains form a subgroup $\mathcal{C}_q(K, G)$ of C^q.

The *coboundary* δc^q of a finite chain $c^q = \sum g_j\sigma_j^q$ is defined as

$$\delta c^q = \sum_i \left(\sum_j [\sigma_i^{q+1}:\sigma_j^q] g_j\right)\sigma_i^{q+1}.$$

Because of (26.1) δc^q is a finite $(q + 1)$-chain, while, because of (26.2), $\delta\delta c^q = 0$. The operation δ is a homomorphic mapping of $\mathcal{C}_q$ into $\mathcal{C}_{q+1}$. The kernel of this homomorphism is a subgroup $\mathcal{Z}_q(K, G)$ of $\mathcal{C}_q$. The chains of $\mathcal{Z}_q$ are called (finite) *cocycles*:

$$\mathcal{Z}_q(K, G) = [\text{all finite } q\text{-chains } c^q \text{ with } \delta c^q = 0].$$

A *coboundary* is a q-chain of the form δd^{q-1} for some $d^{q-1} \epsilon \mathcal{C}_{q-1}$; these coboundaries form a subgroup

$$\mathcal{B}_q(K, G) = [\text{all finite chains } \delta d^{q-1}].$$

From the relation $\delta\delta = 0$ it follows that $\mathcal{B}_q \subset \mathcal{Z}_q$. The corresponding factor group

$$\mathcal{H}_q(K, G) = \mathcal{Z}_q(K, G)/\mathcal{B}_q(K, G)$$

is called the q^{th} *cohomology* group of finite cocycles of K with coefficients in G. We also define the *co-torsion* group $\mathcal{T}_q(K, G)$ as the subgroup of all elements of finite order in $\mathcal{H}_q(K, G)$.

For a chain $c^q = \sum g_i\sigma_i^q$ of $C^q(K, G)$ we also define the *boundary*

$$\partial c^q = \sum_j \left(\sum_i [\sigma_i^q : \sigma_j^{q-1}] g_i\right)\sigma_j^{q-1}.$$

It again follows from (26.1) that ∂c^q is a well defined chain of $C^{q-1}(K, G)$ and from (26.2) that $\partial\partial c^q = 0$. The operation ∂ is a homomorphic mapping of C^q into C^{q-1}. The kernel of this homomorphism is a subgroup $Z^q(K, G)$ of C^q. The chains of Z^q are called *cycles*:

$$Z^q(K, G) = [\text{all chains } c^q \text{ with } \partial c^q = 0].$$

The chains of the form ∂d^{q+1} where $d^{q+1} \in C^{q+1}$ are the *boundaries*. They form a subgroup

$$B^q(K, G) = [\text{all chains } c^q = \partial d^{q+1}].$$

Because $\partial\partial = 0$ it follows that $B^q \subset Z^q$. The group

$$H^q(K, G) = Z^q(K, G)/B^q(K, G)$$

is called the q^{th} *homology* group of K with coefficients in G.

Let L be a (closed or open) subcomplex of K. Each chain c^q in K, considered as a function on the q-cells, defines a corresponding chain c_L^q in L. If $c^q = \sum g_i\sigma_i^q$, $c_L^q = \sum' g_i\sigma_i^q$ is the sum found by deleting all terms $g_i\sigma_i^q$ for which σ_i^q is not in L. If L is open, then $\partial_L(c_L^q) = (\partial c^q)_L$, so that one can establish the following facts.

LEMMA 27.1. *$c^q \in Z^q(K, G)$ if and only if $c_L^q \in Z^q(L, G)$ for every q-finite open subcomplex L of K.*

LEMMA 27.2. *If $c^q \in B^q(K, G)$ then $c_L^q \in B^q(L, G)$, provided L is an open subcomplex of K.*

A statement analogous to Lemma 27.1 concerning B^q is not generally true. In this connection we define the group $B_w^q(K, G)$ of the *weak boundaries* as follows: $c^q \in B_w^q(K, G)$ provided $c_L^q \in B^q(L, G)$ for every q-finite open subcomplex L of K. For each such open subcomplex L we can construct a subcomplex L' consisting of all q-cells of L, all those $(q + 1)$-cells of L which lie on coboundaries of q-cells of L, and all $(q + i)$-cells of K, for $i > 1$. This subcomplex L' is open, is both q and $(q + 1)$-finite, and has $B^q(L, G) = B^q(L', G)$. Hence we conclude that $c^q \in B_w^q(K, G)$ if and only if $c_L^q \in B^q(L, G)$ for every open subcomplex L of K which is both q- and $(q + 1)$-finite. Clearly $B^q = B_w^q$ when K itself is q-finite.

It follows from Lemmas 27.1 and 27.2 that

$$B^q(K, G) \subset B_w^q(K, G) \subset Z^q(K, G).$$

The factor group

$$H^q_w(K, G) = Z^q(K, G)/B^q_w(K, G)$$

will be called the *weak* q^{th} *homology group* of K with coefficients in G. Clearly $H^q = H^q_w$ when K is q-finite

LEMMA 27.3. *$c^q \in B^q_w(K, G)$ if and only if for each finite subset M of K there is a chain c^q_1 in $K - M$ such that $c^q - c^q_1 \in B^q(K, G)$.*

PROOF. Suppose that $c^q \in B^q_w$. Given the finite set M there is a q-finite open subcomplex L containing M. Since $c^q_L \in B^q(L, G)$ there is a d^{q+1} in L such that $(\partial d^{q+1})_L = c^q_L$. Set $c^q_1 = c^q - \partial d^{q+1}$. Clearly $c^q - c^q_1 \in B^q$ and $(c^q_1)_L = c^q_L - (\partial d^{q+1})_L = 0$, hence $c^q_1 \subset K - L \subset K - M$.

Suppose now that c^q satisfies the condition of Lemma 27.3. Given a q-finite open subcomplex L of K there is a c^q_1 in $K - L$ such that $c^q - c^q_1 \in B^q(K, G)$. There is then a d^{q+1} such that $\partial d^{q+1} = c^q - c^q_1$. Since L is open we have

$$\partial_L(d^{q+1}_L) = (\partial d^{q+1})_L = c^q_L - (c^q_1)_L = c^q_L ;$$

therefore $c^q_L \in B^q(L, G)$.

28. Topology in the homology groups

The group of q-chains $C^q(K, G)$ is isomorphic with $\prod_i G_i$, where $G = G_i$ and the set of indices i is in a 1-1 correspondence with the set of q-cells σ^q_i. Hence, if G is a generalized topological group, we can consider $C^q(K, G)$ as a generalized topological group, under the direct product topology, as defined in §1. If G is topological or compact, then $C^q(K, G)$ is also topological or compact, as the case may be.

The boundary operator ∂, regarded as a homomorphism of C^q into C^{q-1}, is continuous. Since Z^q is the group mapped into 0 by ∂, we obtain

LEMMA 28.1. *If G is topological then $Z^q(K, G)$ is a closed subgroup of $C^q(K, G)$.*

From Lemma 27.3 we deduce

LEMMA 28.2. $B^q(K, G) \subset B^q_w(K, G) \subset \bar{B}^q(K, G)$.

The homology groups $H^q = Z^q/B^q$ and $H^q_w = Z^q/B^q_w$ as factor groups of generalized topological groups are generalized topological groups; this is the way they will be considered in the rest of this paper. Even in the case when G and consequently Z^q is topological the groups H^q and H^q_w may be only generalized topological groups, for B^q and B^q_w need not be closed subgroups of Z^q.

If G is compact and topological, then $Z^q(K, G)$ and $C^{q+1}(K, G)$ are compact; since $B^q(K, G)$ is a continuous image of C^{q+1} (under the operation ∂), $B^q(K, G)$ is compact and therefore closed (see Lemma 1.1). Consequently we obtain

LEMMA 28.3. *If G is compact and topological, then $B^q(K, G) = B^q_w(K, G) = \bar{B}^q(K, G)$, $H^q(K, G) = H^q_w(K, G)$ and the groups are all compact and topological.*

Despite the fact that $\mathfrak{C}_q$ is a subgroup of the generalized topological group C^q we consider $\mathfrak{C}_q$ discrete and consequently the cohomology groups $H_q(K, G)$ are taken discrete.

29. The Kronecker index

Let G be a generalized topological group, H a discrete group and assume that a product $\phi(g, h) \in J$ is given pairing G and H to a group J (see §13).

Given two chains

$$c^q \in C^q(K, G), \qquad d^q \in \mathcal{C}_q(K, H),$$

we define the Kronecker index as

$$c^q \cdot d^q = \sum_i \phi(g_i, h_i) \in J;$$

the summation is finite since d^q is a finite chain. We verify at once that in this way the groups $C^q(K, G)$ and $\mathcal{C}_q(K, H)$ are paired to J.

Given $c^{q+1} \in C^{q+1}(K, G)$ and $d^q \in \mathcal{C}_q(K, H)$ we have

$$(\partial c^{q+1}) \cdot d^q = c^{q+1} \cdot (\delta d^q). \tag{29.1}$$

This is a restatement of the associative law for matrix multiplication, since the operator ∂ is essentially a postmultiplication by the incidence matrix, while the coboundary operator δ is a premultiplication by the same matrix.

We now examine the annihilators relative to the Kronecker index.

$$\mathcal{Z}_q(K, H) \subset \text{Annih } B^q_w(K, G) \subset \text{Annih } B^q(K, G). \tag{29.2}$$

$$Z^q(K, G) \subset \text{Annih } \mathcal{B}_q(K, H). \tag{29.3}$$

Proof. Let $z^q \in B^q_w$ and $w^q \in \mathcal{Z}_q$. Since w^q is finite there is a finite subset M of K such that $w^q \subset M$. In view of Lemma 27.3 there is a cycle $z^q_1 \subset K - M$ and a chain $c^{q+1} \in C^{q+1}(K, G)$ such that $\partial c^{q+1} = z^q - z^q_1$. Consequently

$$z^q \cdot w^q = (z^q - z^q_1) \cdot w^q = (\partial c^{q+1}) \cdot w^q = c^{q+1} \cdot \delta w^q = c^{q+1} \cdot 0 = 0.$$

Therefore $\mathcal{Z}_q \subset \text{Annih } B^q_w$. The proof of (29.3) is analogous.

It follows from (29.2) and (29.3) that

(29.4) *$H^q(K, G)$ and $\mathcal{H}_q(K, H)$ are paired to J,*

(29.5) *$H^q_w(K, G)$ and $\mathcal{H}_q(K, H)$ are paired to J.*

Lemma 29.1. *If G and H are dually paired to J then, relative to the Kronecker index,*

(29.6) *$C^q(K, G)$ and $\mathcal{C}_q(K, H)$ are dually paired to J,*

$$\mathcal{Z}_q(K, H) = \text{Annih } B^q_w(K, G) = \text{Annih } B^q(K, G), \tag{29.7}$$

$$Z^q(K, G) = \text{Annih } \mathcal{B}_q(K, H). \tag{29.8}$$

Proof. Given $c^q = \sum g_i \sigma^q_i \neq 0$ in C^q, we have $g_{i_0} \neq 0$ for some i_0. Select $h \in H$ so that $\phi(g_{i_0}, h) \neq 0$. Consider the chain $d^q = h\sigma^q_{i_0}$. Then $c^q \cdot d^q = \phi(g_{i_0}, h) \neq 0$. This proves that Annih $\mathcal{C}_q(K, H) = 0$. Similarly we prove that Annih $C^q(K, G) = 0$. This establishes (29.6).

Let $d^q \in$ Annih $B^q(K, G)$. Hence $c^{q+1} \cdot (\delta d^q) = (\partial c^{q+1}) \cdot d^q = 0$ for every c^{q+1}, and therefore $\delta d^q = 0$, in view of (29.6). This shows that Annih $B^q \subset \mathfrak{Z}_q$, which, together with (29.2), gives (29.7).

The proof of (29.8) is analogous to the previous one.

We remark that even when the pairing of the coefficient groups G and H is dual, the pairing (29.4) or (29.5) of the homology and cohomology groups need not be dual, as observed by Whitney ([14], p. 42).

We shall be especially interested in the pairing of G with the group I of integers to G by means of the product $\phi(g, m) = mg$. This pairing has the property that Annih $I = 0$. This is half of the definition of a dual pairing; the other half (Annih $G = 0$) may fail in case the order of every element in G divides a fixed integer m. Nevertheless the argument for Lemma 29.1 shows in this case that

$$\text{Annih } \mathfrak{C}_q(K, I) = 0, \tag{29.6'}$$

$$Z^q(K, G) = \text{Annih } \mathfrak{B}_q(K, I). \tag{29.8'}$$

We now introduce a subgroup of the group of cycles by the following definition:

$$A^q(K, G) = \text{Annih } \mathfrak{Z}_q(K, I); \tag{29.9}$$

in other words, $c^q \in A^q(K, G)$ if and only if $c^q \cdot w^q = 0$ for every finite integral cocycle w^q. The position of this group A^q may be described as follows:

$$B^q_w(K, G) \subset A^q(K, G) \subset Z^q(K, G). \tag{29.10}$$

By (29.2) we have $\mathfrak{Z}_q \subset$ Annih B^q_w; consequently $B^q_w \subset$ Annih $\mathfrak{Z}_q = A^q$. Since $\mathfrak{B}_q \subset \mathfrak{Z}_q$, we have $A^q =$ Annih $\mathfrak{Z}_q \subset$ Annih $\mathfrak{B}_q = Z^q$ by (29.8').

LEMMA 29.2. *If G is a topological group, $A^q(K, G)$ is closed.*

This follows immediately from the continuity of the Kronecker index.

In case G is topological, the various subgroups of cycles of $C^q(K, G)$ are therefore related as follows:

$$B^q \subset B^q_w \subset \bar{B}^q \subset A^q = \bar{A}^q \subset Z^q = \bar{Z}^q \subset C^q.$$

30. Construction of homomorphisms

The essential device of this chapter is that of using the Kronecker index to generate homomorphisms. For a given chain $c^q \in C^q(K, G)$ define θ_{c^q} by

$$\theta_{c^q}(d^q) = c^q \cdot d^q, \qquad d^q \in \mathfrak{C}_q(K, I). \tag{30.1}$$

LEMMA 30.1. *The correspondence $c^q \to \theta_{c^q}$ establishes an isomorphism*

$$C^q(K, G) \cong \text{Hom } \{\mathfrak{C}_q(K, I), G\}.$$

PROOF. It is clear that $\theta_{c^q} \in$ Hom $\{\mathfrak{C}_q, G\}$, and that the correspondence $c^q \to \theta_{d^q}$ preserves sums. Also, if $\theta_{c^q} = 0$ then $c^q \cdot d^q = 0$ for all $d^q \in \mathfrak{C}_q$ and consequently $c^q = 0$. Conversely, given $\theta \in$ Hom $\{\mathfrak{C}_q, G\}$, define

$$c^q = \sum_i \theta(\sigma^q_i)\sigma^q_i. \tag{30.2}$$

Clearly $c^q \in C^q(K, G)$, while, for any given $d^q = \sum h_i\sigma_i^q \in \mathfrak{C}_q$, we have

$$\theta_{c^q}(d^q) = c^q \cdot d^q = \sum h_i\theta(\sigma_i^q) = \theta(\sum h_i\sigma_i^q) = \theta(d^q).$$

This establishes the algebraic part of the Lemma.

We now recall that

$$C^q(K, G) \cong \prod_i G_i$$

where $G_i = G$ and the subscripts i are in a 1-1 correspondence with the q-cells σ_i^q. On the other hand, since the $\{\sigma_i^q\}$ constitute a set of generators for $\mathfrak{C}_q(K, I)$, we have

$$\text{Hom } \{\mathfrak{C}_q(K, I), G\} \cong \prod_i G_i.$$

Both these isomorphisms are bicontinuous, hence the combined isomorphism, which is precisely the isomorphism $c^q \leftrightarrow \theta_{c^q}$, is also bicontinuous.

LEMMA 30.2. *$\mathfrak{Z}_q(K, I)$ is a direct factor of $\mathfrak{C}_q(K, I)$.*

PROOF. The coboundary operator δ maps $\mathfrak{C}_q$ onto $\mathfrak{B}_{q+1}$ and the kernel is $\mathfrak{Z}_q$. Hence $\mathfrak{C}_q$ is a group extension of $\mathfrak{Z}_q$ by $\mathfrak{B}_{q+1}$. As a subgroup of the free group $\mathfrak{C}_{q+1}$ the group $\mathfrak{B}_{q+1}$ is free (Lemma 4.1) and therefore the group extension is trivial (Theorem 7.2). Hence $\mathfrak{C}_q$ is the direct product of $\mathfrak{Z}_q$ and a subgroup isomorphic with $\mathfrak{B}_{q+1}$.

THEOREM 30.3. *$A^q(K, G)$ is a direct factor of $Z^q(K, G)$ and of $C^q(K, G)$.*

PROOF. Since $A^q \subset Z^q$ it will be sufficient to show that A^q is a direct factor of C^q. In the group Hom $\{\mathfrak{C}_q(K, I), G\}$ consider the subgroup A of those homomorphisms that annihilate Z_q. Since $\mathfrak{Z}_q$ is a direct factor of $\mathfrak{C}_q$, A is a direct factor of Hom $\{\mathfrak{C}_q, G\}$. However, under the isomorphism $\theta_{c^q} \to c^q$ of Lemma 30.1 the group A is mapped onto $A^q(K, G) = \text{Annih } \mathfrak{Z}_q$, hence the conclusion. This proof also shows (Lemma 3.3) that

$$A^q(K, G) \cong \text{Hom } \{\mathfrak{C}_q(K, I)/\mathfrak{Z}_q(K, I), G\}. \tag{30.3}$$

Theorem 30.3 leads to the following direct product decompositions of the homology groups:

$$H^q(K, G) \cong (Z^q/A^q) \times (A^q/B^q), \tag{30.4}$$

$$H_w^q(K, G) \cong (Z^q/A^q) \times (A^q/B_w^q). \tag{30.5}$$

We proceed with the study of the first factor, Z^q/A^q.

THEOREM 30.4. *The correspondence $c^q \to \theta_{c^q}$ establishes an isomorphism*

$$Z^q(K, G)/A^q(K, G) \cong \text{Hom } \{\mathfrak{H}_q(K, I), G\}.$$

PROOF. Since $Z^q = \text{Annih } \mathfrak{B}_q$, by (29.8′), it follows that under the isomorphism $c^q \to \theta_{c^q}$ the group Z^q is mapped onto the subgroup of Hom $\{\mathfrak{C}_q, G\}$ consisting of those homomorphisms annihilating $\mathfrak{B}_q$. By Lemma 3.3 the latter subgroup can be identified with Hom $\{\mathfrak{C}_q/\mathfrak{B}_q, G\}$, so $Z^q \cong \text{Hom } \{\mathfrak{C}_q/\mathfrak{B}_q, G\}$. On the other hand, $\mathfrak{Z}_q/\mathfrak{B}_q$ is a direct factor of $\mathfrak{C}_q/\mathfrak{B}_q$, so that Lemma 3.4 shows

that Hom $\{\mathfrak{Z}_q/\mathfrak{B}_q, G\}$ is a factor group of Hom $\{\mathfrak{C}_q/\mathfrak{B}_q, G\}$, corresponding to the subgroup consisting of homomorphisms annihilating $\mathfrak{Z}_q/\mathfrak{B}_q$. This subgroup in turn corresponds to the subgroup A^q of Z^q, hence

$$Z^q/A^q \cong \text{Hom } \{\mathfrak{Z}_q/\mathfrak{B}_q, G\}.$$

This is the desired conclusion.

31. Study of A^q

The correspondence $c^q \to \theta_{c^q}$ of Lemma 30.1 maps the group A^q of annihilators of cocycles onto the group of those homomorphisms of $\mathfrak{C}_q$ into G which carry $\mathfrak{Z}_q$ into zero. As observed in Lemma 3.3, the latter group is isomorphic to Hom $\{\mathfrak{C}_q/\mathfrak{Z}_q, G\}$. Since $\mathfrak{C}_q/\mathfrak{Z}_q \cong \mathfrak{B}_{q+1}$, this gives the isomorphism

$$(31.1) \qquad A^q(K, G) \cong \text{Hom } \{\mathfrak{B}_{q+1}(K, I), G\}.$$

An examination of this construction shows that the homomorphism corresponding to a given $z^q \in A^q$ is determined as follows. For each $d^{q+1} \in \mathfrak{B}_{q+1}$ choose a $d^q \in \mathfrak{C}_q(K, I)$ for which $\delta d^q = d^{q+1}$, and define[27]

$$\phi_{z^q}(d^{q+1}) = z^q \cdot d^q.$$

Because z^q is in A^q, this result is independent of the choice of d^q for given d^{q+1}. Furthermore ϕ_{z^q} is a homomorphism of $\mathfrak{B}_{q+1}$ into G, and it is obtained from θ_{z^q} by the process indicated above, for one has

$$\phi_{z^q}(\delta d^q) = \theta_{z^q}(d^q).$$

We therefore have the following result.

LEMMA 31.1. *The correspondence $z^q \to \phi_{z^q}$ establishes the (bicontinuous) isomorphism* (31.1).

The properties of this isomorphism can be collected in the following

THEOREM 31.2. *The isomorphism $z^q \to \phi_{z^q}$ induces the isomorphisms*

$$A^q(K, G)/B^q(K, G) \cong \text{Hom } \{\mathfrak{B}_{q+1}, G\}/\text{Hom } \{\mathfrak{Z}_{q+1} \mid \mathfrak{B}_{q+1}, G\},$$

$$B^q_w(K, G)/B^q(K, G) \cong \text{Hom}_f \{\mathfrak{B}_{q+1}, G; \mathfrak{Z}_{q+1}\}/\text{Hom } \{\mathfrak{Z}_{q+1} \mid \mathfrak{B}_{q+1}, G\},$$

$$A^q(K, G)/B^q_w(K, G) \cong \text{Hom } \{\mathfrak{B}_{q+1}, G\}/\text{Hom}_f \{\mathfrak{B}_{q+1}, G; \mathfrak{Z}_{q+1}\},$$

where $\mathfrak{B}_{q+1} = \mathfrak{B}_{q+1}(K, I)$ *and* $\mathfrak{Z}_{q+1} = \mathfrak{Z}_{q+1}(K, I)$.

PROOF. We shall show that the groups $B^q(K, G)$ and $B^q_w(K, G)$ are mapped onto Hom $\{\mathfrak{Z}_{q+1} \mid \mathfrak{B}_{q+1}, G\}$ and Hom$_f$ $\{\mathfrak{B}_{q+1}, G; \mathfrak{Z}_{q+1}\}$, respectively.

Assume that $z^q \in B^q(K, G)$; then $\partial z^{q+1} = z^q$ for some $z^{q+1} \in C^{q+1}(K, G)$. Define

$$\phi^*(d^{q+1}) = z^{q+1} \cdot d^{q+1}; \qquad d^{q+1} \in \mathfrak{C}_{q+1}.$$

---

[27] Notice the analogy with the definition of the so-called "linking coefficient" (cf. Lefschetz [7], Ch. III).

Clearly $\phi^* \in \text{Hom}\,\{\mathcal{C}_{q+1}, G\}$. If $d^{q+1} = \delta d^q$ then

$$\phi^*(d^{q+1}) = z^{q+1}\cdot d^{q+1} = z^{q+1}\cdot\delta d^q = \partial z^{q+1}\cdot d^q = z^q\cdot d^q = \phi_{z^q}(d^{q+1}).$$

Hence ϕ^* is an extension of ϕ_{z^q} to $\mathcal{C}_{q+1}$ and in particular also to $\mathcal{Z}_{q+1}$.

Suppose conversely that ϕ_{z^q} can be extended to $\mathcal{Z}_{q+1}$. Since $\mathcal{Z}_{q+1}$ is a direct factor of $\mathcal{C}_{q+1}$ (Lemma 30.2) we may then find an extension ϕ^* of ϕ_{z^q} to $\mathcal{C}_{q+1}$. Define

$$z^{q+1} = \sum_i \phi^*(\sigma_i^{q+1})\sigma_i^{q+1}.$$

Clearly $z^{q+1} \in C^{q+1}(K, G)$ and $z^{q+1}\cdot\sigma_j^{q+1} = \phi^*(\sigma_j^{q+1})$ and hence $z^{q+1}\cdot d^{q+1} = \phi^*(d^{q+1})$ for all $d^{q+1} \in \mathcal{C}_{q+1}$. Consequently

$$\partial z^{q+1}\cdot\sigma_j^q = z^{q+1}\cdot\delta\sigma_j^q = \phi^*(\delta\sigma_j^q) = \phi_{z^q}(\delta\sigma_j^q) = z^q\cdot\sigma_j^q.$$

Since this holds for every σ_j^q we have $\partial z^{q+1} = z^q \in B^q(K, G)$.

Suppose $z^q \in B_w^q(K, G)$. In view of Lemma 5.1 it is sufficient to prove that if the cocycle $md^{q+1} \in \mathcal{B}_{q+1}(K)$ then $\phi_{z^q}(md^{q+1})$ is divisible by m. Let $\delta d^q = md^{q+1}$ and let M be a finite subset of K such that $d^q \subset M$. In view of Lemma 27.3 there is a chain $z_1^q \subset K - M$ such that $z^q - z_1^q = z_2^q \in B^q(K, G)$. It follows that $z_2^q \in A^q$ and so that $z_1^q \in A^q$, hence $\phi_{z_1^q}$ and $\phi_{z_2^q}$ are defined and $\phi_{z^q} = \phi_{z_1^q} + \phi_{z_2^q}$. Since $z_2^q \in B^q(K, G)$, then, as we just proved, $\phi_{z_2^q}$ can be extended to Z_{q+1} and therefore $\phi_{z_2^q}(md^{q+1})$ must be divisible by m. Since $d^q \subset M$ and $z_1^q \subset K - M$ we have $\phi_{z_1^q}(md^{q+1}) = z_1^q\cdot d^q = 0$. Hence $\phi_{z^q}(md^{q+1})$ is divisible by m.

Suppose conversely that ϕ_{z^q} can be extended to every subgroup of $\mathcal{Z}_{q+1}(K, I)$ of finite order over $\mathcal{B}_{q+1}(K, I)$. Then, as in Lemma 5.2, ϕ_{z^q} can also be extended to every subgroup $\mathfrak{D}$ of $\mathcal{Z}_{q+1}(K, I)$ such that $\mathfrak{D}/\mathcal{B}_{q+1}$ has a finite number of generators. Now let L be any open subcomplex of K which is both q and $(q + 1)$-finite; there is then an extension of ϕ_{z^q} to the group $\mathfrak{D}_L$ generated by $\mathcal{B}_{q+1}(K, I)$ and $\mathcal{Z}_{q+1}(L, I)$. But in the complex L the homomorphism ϕ_{y^q} induced by $y^q = z_L^q$ agrees on $\mathcal{B}_{q+1}(L, I)$ with the homomorphism ϕ_{z^q}. Therefore $\phi_{y^q} \in \text{Hom}\,\{\mathcal{B}_{q+1}(L, I), G\}$ has an extension to $\mathcal{Z}_{q+1}(L, I)$. In view of what we proved before, we therefore have $y^q = z_L^q \in B^q(L, G)$. Since this holds for each L considered, $z^q \in B_w^q(K, G)$. This concludes the proof of Theorem 31.2.

In this theorem the factor homomorphism groups on the right can be reinterpreted as groups of group extensions, in accord with the results of Chapter II.

THEOREM 31.3. *The isomorphism $z^q \leftrightarrow \phi_{z^q}$ combined with the isomorphisms establishing relations between group extensions and homomorphisms lead to the following isomorphisms*:

$$A^q(K, G)/B^q(K, G) \cong \text{Ext}\,\{G, \mathcal{H}_{q+1}\}, \tag{31.2}$$

$$B_w^q(K, G)/B^q(K, G) \cong \text{Ext}_f\,\{G, \mathcal{H}_{q+1}\}, \tag{31.3}$$

$$A^q(K, G)/B_w^q(K, G) \cong \text{Ext}\,\{G, \mathcal{T}_{q+1}\}/\text{Ext}_f\,\{G, \mathcal{T}_{q+1}\} \tag{31.4}$$

where $\mathcal{H}_{q+1} = \mathcal{H}_{q+1}(K, I)$ and $\mathcal{T}_{q+1} = \mathcal{T}_{q+1}(K, I)$ is the corresponding co-torsion group.

The isomorphisms established so far have all been bicontinuous.

32. Computation of the homology groups

As we have shown in §29, the Kronecker index establishes a pairing of the group $H^q(K, G)$ or $H^q_w(K, G)$ with the group $\mathcal{H}_q(K, I)$, the values of the products being in the group G. Accordingly we define the following subhomology groups:

(32.1) $$Q^q(K, G) = \text{Annih } \mathcal{H}^q(K, I) \text{ in } H^q(K, G),$$

(32.2) $$Q^q_w(K, G) = \text{Annih } \mathcal{H}_q(K, I) \text{ in } H^q_w(K, G).$$

We verify at once that $Q^q = A^q/B^q$ and $Q^q_w = A^q/B^q_w$. Consequently the results of the last two sections furnish the following two basic theorems:

THEOREM 32.1. *For a star finite complex K the homology group $H^q(K, G)$ of infinite cycles with coefficients in a generalized topological group G can be expressed in terms of the integral cohomology groups $\mathcal{H}_q = \mathcal{H}_q(K, I)$ and $\mathcal{H}_{q+1} = \mathcal{H}_{q+1}(K, I)$ of finite cocycles. The explicit relation is*

(32.3) $$H^q(K, G) \cong \text{Hom } \{\mathcal{H}_q, G\} \times \text{Ext } \{G, \mathcal{H}_{q+1}\}.$$

More explicitly, H^q has a subgroup Q^q, defined by (32.1), *where*

(32.4) $$Q^q(K, G) \textit{ is a direct factor of } H^q(K, G),$$

(32.5) $$Q^q(K, G) \cong \text{Ext } \{G, \mathcal{H}_{q+1}\},$$

(32.6) $$H^q(K, G)/Q^q(K, G) \cong \text{Hom } \{\mathcal{H}_q, G\}.$$

THEOREM 32.2. *For a star finite complex K the weak homology group $H^q_w(K, G)$ of infinite cycles with coefficients in a generalized topological group G can be expressed in terms of the integral cohomology group $\mathcal{H}_q = \mathcal{H}_q(K, I)$ and the integral co-torsion group $\mathcal{T}_{q+1} = \mathcal{T}_{q+1}(K, I)$ of finite cocycles. The explicit relation is*

(32.7) $$H^q_w(K, G) \cong \text{Hom } \{\mathcal{H}_q, G\} \times (\text{Ext } \{G, \mathcal{T}_{q+1}\}/\text{Ext}_f \{G, \mathcal{T}_{q+1}\}).$$

More explicitly, H^q_w has a subgroup Q^q_w, defined by (32.2), *where*

(32.8) $$Q^q_w(K, G) \textit{ is a direct factor of } H^q_w(K, G),$$

(32.9) $$Q^q_w(K, G) \cong \text{Ext } \{G, \mathcal{T}_{q+1}\}/\text{Ext}_f \{G, \mathcal{T}_{q+1}\},$$

(32.10) $$H^q_w(K, G)/Q^q_w(K, G) \cong \text{Hom } \{\mathcal{H}_q, G\}.$$

Both factors in (32.3) and (32.7) are generalized topological groups and the isomorphisms are bicontinuous.

If G is topological then by Corollary 3.2 the group Hom $\{\mathcal{H}_q, G\}$ is topological. If we also assume that mG is a closed subgroup of G for $m = 2, 3, \cdots$ then Corollary 11.6 shows that $\text{Ext}_f \{G, \mathcal{T}_{q+1}\}$ is a closed subgroup of Ext $\{G, \mathcal{T}_{q+1}\}$. Consequently we obtain

THEOREM 32.3. (Steenrod [9]). *If G is a topological group and mG is a closed subgroup of G for $m = 2, 3, \cdots$ then $H^q_w(K, G)$ is topological.*

The expressions for Q^q and Q^q_w can be simplified if additional information concerning the group G is available. If G is infinitely divisible then, by Corollary 11.4, Ext $\{G, H\} = 0$ for all H and therefore

COROLLARY 32.4. *If G is infinitely divisible then* $Q^q(K, G) = Q^q_w(K, G) = 0$ *and*

$$H^q(K, G) = H^q_w(K, G) \cong \text{Hom}\ \{\mathcal{H}_q, G\}.$$

From Theorem 17.2 we deduce

COROLLARY 32.5. *If G has no elements of finite order then*

$$Q^q_w(K, G) \cong \text{Ext}\ \{G, \mathcal{T}_{q+1}\}.$$

If, in addition, G is discrete then

$$Q^q_w(K, G) \cong \text{Hom}\ \{\mathcal{T}_{q+1}, G_\infty/G\}.$$

In particular, if $G = I$ then, by Theorem 17.1, $Q^q_w(K, I) \cong \text{Char}\ \mathcal{T}_{q+1}$ and therefore

$$H^q_w(K, I) \cong \text{Hom}\ \{\mathcal{H}_q, I\} \times \text{Char}\ \mathcal{T}_{q+1}. \tag{32.11}$$

THEOREM 32.6. *If G is compact and topological then* $H^q(K, G) = H^q_w(K, G)$ *is compact and topological and*

$$Q^q(K, G) = Q^q_w(K, G) \cong \text{Ext}\ \{G, \mathcal{T}_{q+1}\} \cong \text{Char Hom}\ \{G, \mathcal{T}_{q+1}\}. \tag{32.12}$$

This is a consequence of Corollary 11.7 and Theorem 15.1. Since G is compact, $\mathcal{T}_{q+1}$ discrete, and only continuous homomorphisms are taken in Hom $\{G, \mathcal{T}_{q+1}\}$, it follows that in the formula (32.12) for $Q^q(K, G)$ we may replace G by G/G_0 where G_0 is the component of 0 in G.

COROLLARY 32.7. *If* $\mathcal{H}_{q+1}(K, I)$ *has a finite number of generators then* $B^q(K, G) = B^q_w(K, G)$ *and*

$$H^q(K, G) = H^q_w(K, G) \cong \text{Hom}\ \{\mathcal{H}_q, G\} \times \text{Ext}\ \{G, \mathcal{T}_{q+1}\}. \tag{32.13}$$

In fact, since $\text{Ext}_f\ \{G, \mathcal{H}_{q+1}\} = 0$ (Corollary 11.3) it follows from (31.3) that $B^q = B^q_w$. Since also $\text{Ext}_f\ \{G, \mathcal{T}_{q+1}\} = 0$, formula (32.13) follows from Theorem 32.2.

In particular, Corollary 32.7 applies if K is a finite complex (cf. Alexandroff-Hopf [1], Ch. V and Steenrod [9], p. 675).

33. Computation of the cohomology groups

We start out with a brief review of the duality between homology and cohomology. Let G be a discrete group and $\hat{G} = \text{Char}\ G$ compact and topological. Since $\hat{G}$ and G are dually paired to the group P of reals mod 1 (see §13) the Kronecker index $c^q \cdot d^q \in P$ is defined as in §29 for $c^q \in C^q(K, \hat{G})$ and $d^q \in \mathcal{C}_q(K, G)$. Since the pairing of $\hat{G}$ and G is dual (Theorem 13.5) we have by Lemma 29.1

(33.1) $C^q(K, \hat{G})$ *and* $\mathcal{C}_q(K, G)$ *are dually paired to P,*

$$\mathcal{Z}_q(K, G) = \text{Annih}\ B^q(K, \hat{G}); \qquad Z^q(K, \hat{G}) = \text{Annih}\ \mathcal{B}_q(K, G). \tag{33.2}$$

These formulas, Theorem 13.7, Lemma 13.2 and Theorem 13.5 imply that the Kronecker index defines a dual pairing of $\mathcal{H}_q(K, G)$ and $H^q(K, \hat{G})$ to P and that

$$\mathcal{H}_q(K, G) \cong \text{Char}\ H^q(K, \text{Char}\ G). \tag{33.3}$$

Using this result and the formulas established in the previous section for $H^q(K, \text{Char } G)$ we could write down a formula expressing $\mathcal{H}_q(K, G)$. For convenience we first define a subcohomology group

$$(33.4) \qquad \mathcal{P}_q(K, G) = \text{Annih } Q^q(K, \text{Char } G) \text{ in } \mathcal{H}_q(K, G),$$

in order to get a more detailed form for our result.

THEOREM 33.1. *For a star finite complex K the cohomology group $\mathcal{H}_q(K, G)$ of finite cocycles with coefficients in a discrete group G can be expressed in terms of the cohomology group $\mathcal{H}_q = \mathcal{H}_q(K, I)$ and the integral co-torsion group $\mathcal{T}_{q+1} = \mathcal{T}_{q+1}(K, I)$. The explicit relation is*

$$(33.5) \qquad \mathcal{H}_q(K, G) \cong (G \circ \mathcal{H}_q) \times \text{Hom } \{\text{Char } G, \mathcal{T}_{q+1}\}.$$

More explicitly, $\mathcal{H}_q(K, G)$ has a subgroup $\mathcal{P}_q(K, G)$, defined by (33.4), *where*

$$(33.6) \qquad \mathcal{P}_q(K, G) \textit{ is a direct factor of } \mathcal{H}_q(K, G),$$

$$(33.7) \qquad \mathcal{P}_q(K, G) \cong G \circ \mathcal{H}_q$$

$$(33.8) \qquad \mathcal{H}_q(K, G)/\mathcal{P}_q(K, G) \cong \text{Hom } \{\text{Char } G, \mathcal{T}_{q+1}\}.$$

PROOF. Since Q^q is a direct factor of H^q it follows from the character theory that $\mathcal{P}_q = \text{Annih } Q^q$ is a direct factor of $\mathcal{H}_q(K, G) = \text{Char } H^q$. It also follows that

$$\mathcal{P}_q \cong \text{Char } (H^q/Q^q), \qquad \mathcal{H}_q(K, G)/\mathcal{P}_q(K, G) \cong \text{Char } Q^q.$$

The first formula and (32.6) imply

$$\mathcal{P}_q(K, G) \cong \text{Char Hom } \{\mathcal{H}_q, \text{Char } G\},$$

which in view of Theorem 18.1 gives (33.7). The second formula combined with (32.12) proves (33.8).

If G has no elements of finite order, then Char G is connected and therefore Hom $\{\text{Char } G, \mathcal{T}_{q+1}\} = 0$. From (33.7) and (33.8) we therefore obtain

COROLLARY 33.2. *If G has no elements of finite order then*

$$\mathcal{H}_q(K, G) = \mathcal{P}_q(K, G) \cong G \circ \mathcal{H}_q(K, I).$$

We now proceed to give an intrinsic characterization of the subgroup $\mathcal{P}_q$ of $\mathcal{H}_q(K, G)$. A cocycle $w^q \in \mathcal{Z}_q(K, G)$ will be called *pure* if it is a linear combination of integral cocycles, as

$$w_q = \sum_{i=1}^{k} g_i w_i^q, \qquad g_i \in G, \qquad w_i^q \in \mathcal{Z}_q(K, I).$$

LEMMA 33.3. *The group $\mathcal{P}_q(K, G)$ is the subgroup of $\mathcal{H}_q(K, G)$ determined by the pure cocycles.*

PROOF. Let $\mathcal{S}$ be the subgroup of $\mathcal{Z}_q(K, G)$ consisting of all the pure cocycles. It may be shown that $\mathcal{B}_q(K, G) \subset \mathcal{S}$. In order to prove that $\mathcal{S}/\mathcal{B}_q(K, G) = \mathcal{P}_q(K, G)$ we must prove that $\mathcal{S}/\mathcal{B}_q(K, G) = \text{Annih } Q^q(K, \hat{G})$ where $\hat{G} = \text{Char } G$.

This is equivalent to proving that $Q^q(K, \hat{G}) = \text{Annih } (\mathcal{S}/\mathcal{B}_q(K, G))$, which reduces to the formula

$$A^q(K, \hat{G}) = \text{Annih } \mathcal{S},$$

that we now propose to establish.

Let $z^q \in A^q(K, \hat{G})$ and let $w^q \in \mathcal{S}$. Since $w^q = \sum g_i w_i^q$, where $w_i^q \in \mathcal{Z}_q(K, I)$ and since $z^q \cdot w_i^q = 0$ by the definition of A^q, it follows that $z^q \cdot w^q = 0$.

Suppose now that c^q lies in $C^q(K, \hat{G})$ but not in $A^q(K, \hat{G})$. There is then a $w_i^q \in \mathcal{Z}_q(K, I)$ such that $c^q \cdot w_i^q = \hat{g} \neq 0$ where $\hat{g} \in \hat{G}$. Pick $g \in G$ so that $\hat{g}(g) \neq 0$ and define $w^q = g w_1^q$. Clearly $w^q \in \mathcal{S}$ is a pure cocycle and $c^q \cdot w^q = \hat{g}(g) \neq 0$, hence c^q is not in Annih $\mathcal{S}$. This concludes the proof of the Lemma.

Using the description of $\mathcal{P}_q(K, G)$ given in the Lemma we could easily establish the isomorphism $\mathcal{P}_q \cong G \circ \mathcal{H}_q(K, I)$ directly, using the definition of the tensor product. This was the procedure adopted by Čech [3] who essentially has proved all the results of this section. Our main improvement is that our isomorphisms are given explicitly and invariantly, while Čech used generators and relations throughout.

34. The groups H_t^q

The fact that the groups H^q and H_w^q may not be topological groups even though the coefficient group G is chosen to be topological induced Lefschetz and others to introduce the following group, for a topological coefficient group G,

$$H_t^q(K, G) = Z^q(K, G)/\bar{B}^q(K, G)$$

as a standard homology group for K.

The relation of this group to the groups previously considered is immediate:

$$H_t^q \cong H^q/\bar{0} \cong H_w^q/\bar{0}. \tag{34.1}$$

Theorem 32.3 can now be reformulated as follows.

THEOREM 34.1 (Steenrod [9]) *If G is topological and mG is closed for $m = 2, 3, \cdots$ then $H_w^q(K, G) = H_t^q(K, G)$.*

Since G is a topological group, $A^q(K, G)$ is a closed subgroup of $Z^q(K, G)$ (Lemma 29.2) and consequently $\bar{B}^q \subset A^q$. It follows that the Kronecker index can be defined for elements of $H_t^q(K, G)$ and $\mathcal{H}_q(K, I)$. We define a subhomology group

$$Q_t^q(K, G) = \text{Annih } \mathcal{H}_q(K, I) \text{ in } H_t^q(K, G). \tag{34.2}$$

THEOREM 34.2. *For a star finite complex K the topological homology group $H_t^q(K, G)$ of infinite cycles with coefficients in a topological group G can be expressed in terms of the integral cohomology group $\mathcal{H}_q = \mathcal{H}_q(K, I)$ and the integral co-torsion group $\mathcal{T}_{q+1} = \mathcal{T}_{q+1}(K, I)$ of finite cocycles. The explicit relation is*

$$H_t^q(K, G) \cong \text{Hom } \{\mathcal{H}_q, G\} \times (\text{Ext } \{G, \mathcal{T}_{q+1}\}/\bar{0}). \tag{34.3}$$

More explicitly, H_t^q *has a subgroup* Q_t^q, *defined by* (34.2), *where*

(34.3) $Q_t^q(K, G)$ *is a direct factor of* $H_t^q(K, G)$,

(34.5) $Q_t^q(K, G) \cong \text{Ext}\,\{G, \mathcal{T}_{q+1}\}/\bar{0}$,

(34.6) $H_t^q(K, G)/Q_t^q(K, G) \cong \text{Hom}\,\{\mathcal{H}_q, G\}$.

PROOF. From the direct product decomposition (30.5) we obtain

$$H_t^q \cong (Z^q/A^q) \times [(A^q/B_w^q)/\bar{0}].$$

Consequently $Q_t^q = Q_w^q/\bar{0}$ is a direct factor. Since $Q_w^q \cong \text{Ext}\,\{G, \mathcal{T}_{q+1}\}/\text{Ext}_f\,\{G, \mathcal{T}_{q+1}\}$ and since, by Corollary 11.6, $\overline{\text{Ext}_f}\,\{G, \mathcal{T}_{q+1}\} = \bar{0}$, we obtain (34.5). Formula (34.6) follows from Theorem 30.4.

It might be interesting to notice that, while the groups $H^q(K, G)$ and $H_w^q(K, G)$ were algebraically independent of the choice of the topology in G, the group $H_t^q(K, G)$ depends both algebraically and topologically upon the topology chosen in G.

35. Universal coefficients

The results of the previous three sections can be summarized in the following fashion.

UNIVERSAL COEFFICIENT THEOREM. *In a star finite complex K the integral cohomology groups of finite cocycles determine all the homology and cohomology groups that were defined for a star finite complex, specifically:*

The groups G, $\mathcal{H}_q(K, I)$ and $\mathcal{H}_{q+1}(K, I)$ determine the generalized topological homology group $H^q(K, G)$ of infinite cycles with coefficients in a generalized topological group G.

The groups G, $\mathcal{H}_q(K, I)$ and $\mathcal{T}_{q+1}(K, I)$ determine:

(a) *the generalized topological weak homology group $H_w^q(K, G)$ of infinite cycles with coefficients in a generalized topological group G;*

(b) *the topological homology group $H_t^q(K, G)$ of infinite cycles with coefficients in a topological group G;*

(c) *the discrete cohomology group $\mathcal{H}_q(K, G)$ of finite cocycles with coefficients in a discrete group G.*

This shows that the group I of integers is a universal coefficient group for the homology theory of the complex K. Since the group P of reals mod 1 is the group of characters cf I we have in view of (33.3) the fact that $\mathcal{H}_q(K, I) \cong \text{Char}\, H^q(K, P)$; therefore all the groups can be expressed in terms of $H^q(K, P)$ and $H^{q+1}(K, P)$, so that P is also universal.

Given a closed subcomplex L of K one often has to consider the relative groups of K mod L. However, the complexes used here are so general that $K - L$ is also a complex and the usual groups of K mod L coincide with the groups of $K - L$ as we have defined them. Consequently all our formulas remain valid in the relative theory.

36. Closure finite complexes

Closure finite complexes are obtained by replacing condition (26.1) in the definition of a complex by the following

(36.1) *Given* σ_i^q, $[\sigma_i^q : \sigma_k^{q-1}] \neq 0$ *for only a finite number of indices* k.

Simplicial complexes are all closure finite.

In a closure finite complex we consider finite cycles and infinite cocycles and obtain the discrete homology groups $\mathcal{H}^q(K, G)$ and the topologized cohomology groups $H_q(K, G)$, $H_q^w(K, G)$ and $H_q^t(K, G)$. All our development can be repeated with the modification of interchanging homology and cohomology groups and replacing $q + 1$ by $q - 1$. For instance formula (32.3) will take the form:

$$H_q(K, G) \cong \mathrm{Hom}\,\{\mathcal{H}^q(K, I), G\} \times \mathrm{Ext}\,\{G, \mathcal{H}^{q-1}(K, I)\}.$$

Instead of repeating the arguments for closure finite complexes we can use the previous results for star finite complexes and apply them to closure finite complexes by means of the concept of the dual complex. If the complex K is described by the incidence matrices A^q, the dual complex K^* will be defined by the transposed matrices

$$B^q = (A^{-q})'$$

The dual of a star finite complex is closure finite and vice versa. Also $(K^*)^* = K$. Moreover by passing from a complex to its dual, the boundary operation becomes the coboundary, and vice versa. Hence the homology and cohomology group are interchanged, and our formulas apply.

A locally finite (i.e. both closure and star finite) complex carries therefore two homology theories, namely, the theory of a star finite complex and the theory of a closure finite one. In the case of a manifold the Poincaré duality establishes a relation between the two theories. In general the theories are unrelated and in any specific problem we only use one at a time. We will quote two examples to this effect.

A) In the following chapter we define for every compact metric space a complex called the fundamental complex. This complex is locally finite, but its closure finite theory is trivial, while its star finite theory is extremely useful for the study of the underlying space.

B) Let us consider two infinite polyhedra represented as two locally finite complexes K and K'. Given a continuous mapping f of K into K' it is well known that f induces homomorphisms: 1°) of the groups of finite cycles of K into the corresponding groups of K', 2°) of the groups of infinite cocycles of K' into the corresponding groups of K. This explains why in problems connected with continuous mappings (like Hopf's mapping theorem and its generalizations; see [4]) we use only finite cycles and infinite cocycles, or in other words we use only the closure finite theory of K and K'.

CHAPTER VI. TOPOLOGICAL SPACES

Here we formulate our results for the homology groups of a space. In the case of a compact metric space, Steenrod has shown that the homology groups can all be expressed as corresponding homology groups of the fundamental complex of the space, so that the results of Chapter V apply directly (§44). For a general space, the Čech homology groups are obtained as (direct or inverse) limits, so that the decomposition of the homology group is obtained as a limit of the known decompositions for the homology groups of finite complexes, and here the techniques developed in Chapter IV apply. The results obtained for a general space are not as complete as those for complexes, partly because the limit of a set of direct sums apparently need not be a direct sum, and partly because "Lim" and "Ext" do not permute, so that the group Ext* discussed in Chap. IV is requisite. We also discuss (§45) Steenrod's homology groups of "regular" cycles.

37. Chain transformations

Let $K = \{\sigma_i^q\}$ and $K' = \{\tau_j^q\}$ be two star finite complexes. Suppose also that for every integer q there is given a matrix of integers,

$$B^q = \| b_{ij}^q \|$$

with rows indexed by the q-cells of K, columns by the q-cells of K', and with only a finite number of non-zero entries in each column.

Given a q-chain $c^q = \sum g_i \sigma_i^q \in C^q(K, G)$ in K, define

$$Tc^q = \sum_j \left(\sum_i g_i b_{ij}^q\right) \tau_j^q.$$

The column finiteness condition implies that the summation $\sum_i g_i b_{ij}^q$ is finite and therefore that Tc^q is a well defined element of $C^q(K', G)$. We thus obtain homomorphisms (one for each q and G)

$$T\colon \quad C^q(K, G) \to C^q(K', G).$$

Given a finite q-chain $d^q = \sum g_j \tau_j^q \in \mathcal{C}_q(K', G)$ in K', define

$$T^* d^q = \sum_i \left(\sum_j g_j b_{ij}^q\right) \sigma_i^q$$

This time the column finiteness of B^q implies that T^*d^q is finite; hence we obtain homomorphisms

$$T^*\colon \quad \mathcal{C}_q(K', G) \to \mathcal{C}_q(K, G).$$

T^* is called the *dual* of T.

It can be verified at once that if c^q is a chain in K and d^q is a finite chain in K' then

$$(Tc^q) \cdot d^q = c^q \cdot (T^* d^q), \tag{37.1}$$

whenever the coefficients are such that the Kronecker index has a meaning (§29).

T is called a *chain transformation* of K into K' if $\partial Tc^q = T(\partial c^q)$ for every q chain; that is, if

$$\partial T = T\partial. \tag{37.2}$$

It can be shown that this condition is equivalent to the requirement that

$$\delta T^* = T^*\delta. \tag{37.3}$$

It follows that a chain transformation T maps the groups Z^q, A^q, B^q_w and B^q of K homomorphically into the corresponding groups of K'. Similarly T^* maps the groups of K' into the corresponding groups of K. In particular a chain transformation induces homomorphisms of the homology groups

$$T\colon\ H^q(K, G) \to H^q(K', G), \tag{37.4}$$

$$T^*\colon\ \mathcal{H}_q(K', G) \to \mathcal{H}_q(K, G), \tag{37.5}$$

and of the corresponding subgroups defined by (32.1) and (33.4)

$$T\colon\ Q^q(K, G) \to Q^q(K', G), \tag{37.6}$$

$$T^*\colon\ \mathcal{P}_q(K', G) \to \mathcal{P}_q(K, G). \tag{37.7}$$

38. Naturality

We are now in a position to give a precise meaning to the fact that the isomorphisms established in Chapter V are all "natural."

THEOREM 38.1. *If T is a chain transformation of a complex K into K', then T permutes with the isomorphisms established in Theorems* 30.4 *and* 31.2, *provided the application of T in any group is taken to mean the application of the appropriate transformation induced by T on that group.*

PROOF. If the homomorphism established in Theorem 30.4 be denoted by μ (or by μ', for K'), then we have the homomorphisms

$$\begin{tikzcd}
Z^q(K) \arrow[r, "\mu"] \arrow[d, "T"'] & \operatorname{Hom}\{\mathcal{H}_q, G\} \arrow[d, "T_h^{**}"] \\
Z^q(K') \arrow[r, "\mu'"] & \operatorname{Hom}\{\mathcal{H}'_q, G\},
\end{tikzcd}$$

where T_h^{**} is the homomorphism of $\operatorname{Hom}\{\mathcal{H}_q(K, I), G\}$ into $\operatorname{Hom}\{\mathcal{H}_q(K', I), G\}$, induced as in §12 by the dual chain transformation T^*. The theorem then asserts that

$$\mu' T = T_h^{**}\mu.$$

To show this, take $c^q \in Z^q(K, G)$. The corresponding homomorphism $\theta = \mu c^q$ is then defined, for each cocycle d^q in $\mathcal{Z}_q(K)$, by $\theta(d^q) = c^q \cdot d^q$ (cf. §30). Then $\theta' = T_h^{**}\theta$ is, according to the definition of T_h, simply $\theta'(d'^q) = \theta(T^* d'^q)$. Hence, for any cocycle d'^q,

$$\theta'(d'^q) = \theta(T^* d'^q) = c^q \cdot (T^* d'^q) = (Tc^q) \cdot d'^q.$$

In the other direction, Tc^q maps under μ' into the homomorphism ϕ', defined for $d'^q \in Z_q(K')$ by

$$\phi'(d'^q) = (Tc^q) \cdot d'^q.$$

The formulas show that $\phi' = \mu' T c^q$ and $\theta' = T_h^{**} \mu c^q$ are in fact identical, as required by Theorem 38.1.

To treat Theorem 31.2, let τ (or τ') denote the homomorphism of $A^q(K, G)$ onto Hom $\{\mathcal{B}_{q+1}(K, I), G\}$ given in that theorem, while η is the map of the latter group onto Ext $\{G, \mathcal{H}_{q+1}\}$. The figure is

$$\begin{array}{ccccc} A^q & \xrightarrow{\tau} & \text{Hom}\,\{\mathcal{B}_{q+1}, G\} & \xrightarrow{\eta} & \text{Ext}\,\{G, \mathcal{H}_{q+1}\} \\ \downarrow T & & \downarrow T_h^{**} & & \downarrow T_e^{**} \\ A'^q & \xrightarrow[\tau']{} & \text{Hom}\,\{\mathcal{B}'_{q+1}, G\} & \xrightarrow[\eta']{} & \text{Ext}\,\{G, \mathcal{H}'_{q+1}\} \end{array}$$

where T_h^{**}, T_e^{**} are again the induced homomorphisms. If $z^q \in A^q(K, G)$ is given, $\phi = \tau z^q$ is defined on each coboundary δd^q as $\phi(\delta d^q) = z^q \cdot d^q$, while $\phi' = T_h^{**}\phi$ is defined in turn as

$$\phi'(\delta d'^q) = \phi(T^* \delta d'^q) = \phi(\delta T^* d'^q) = z^q \cdot (T^* d'^q).$$

On the other hand, $\chi = \tau'(Tz^q)$ is defined on a coboundary $\delta d'^q$ of K' as

$$\chi(\delta d'^q) = (Tz^q) \cdot d'^q = z^q \cdot (T^* d'^q).$$

The results are identical, so $T_h^{**}\tau = \tau' T$. Now the "naturality" theorem for group extensions showed that T permutes with η, as in $T_e^{**}\eta = \eta' T_h^{**}$. Combination of these results gives

$$(\eta'\tau')T = T_e^{**}(\eta\tau).$$

This is the required commutativity condition, for $\eta\tau$ is the isomorphism envisaged in Theorem 31.3.

39. Čech's homology groups

We now briefly outline Čech's method of defining the homology and cohomology groups for a space X. Let U_α be a finite open covering of X and N_α the nerve of U_α. If U_β is a refinement of U_α we write $\alpha < \beta$. For $\alpha < \beta$ we have a chain transformation $T_{\alpha\beta}$: $N_\beta \rightarrow N_\alpha$ defined as follows: for each open set of the covering U_β select a set of U_α containing it; this maps the vertices of N_β into the vertices of N_α and leads to a simplicial mapping $T_{\alpha\beta}$. This chain transformation is not defined uniquely, but the induced homomorphisms

$$T_{\alpha\beta}: \quad H^q(N_\beta, G) \rightarrow H^q(N_\alpha, G),$$

$$T^*_{\beta\alpha}: \quad \mathcal{H}_q(N_\alpha, G) \rightarrow \mathcal{H}_q(N_\beta, G)$$

are unique. Using the directed system of all the finite open coverings of X we define[28]

$$\mathcal{H}^q(X, G) = \varprojlim H^q(N_\alpha, G) \tag{39.1}$$

$$\mathcal{H}_q(X, G) = \varinjlim \mathcal{H}_q(N_\alpha, G). \tag{39.2}$$

In (39.2) the groups are all discrete. In (39.1) G can be any generalized topological group and $\mathcal{H}^q(X, G)$, as an inverse limit of generalized topological groups, also is a generalized topological group. If G has the property that each of its subgroups mG ($m = 2, 3 \cdots$) is closed in G, the finiteness of each N_α implies that $H^q(N_\alpha, G)$ and hence $\mathcal{H}^q(X, G)$ is topological. If G does not have this property, it would still be possible to consider the group

$$\varprojlim H^q_t(N_\alpha, G) = \varprojlim\, [H^q(N_\alpha, G)/\bar{0}].$$

This group is always topological but its relation to the other groups is rather obscure.

In view of (37.6) the subgroups $Q^q(N_\alpha, G)$ of $H^q(N_\alpha, G)$ form an inverse system. We define

$$\mathcal{Q}^q(X, G) = \varprojlim Q^q(N_\alpha, G). \tag{39.3}$$

Clearly $\mathcal{Q}^q$ is a subgroup of $\mathcal{H}^q(X, G)$.

Similarly, in view of (37.7), the subgroups $\mathcal{P}_q(N_\alpha, G)$ of $\mathcal{H}_q(N_\alpha, G)$ form a direct system so we define

$$\mathcal{P}_q(X, G) = \varinjlim \mathcal{P}_q(N_\alpha, G). \tag{39.4}$$

$\mathcal{P}_q$ is a subgroup of $\mathcal{H}_q(X, G)$.

LEMMA 39.1. *The Kronecker index establishes a pairing of* $\mathcal{H}^q(X, G)$ *and* $\mathcal{H}_q(X, I)$ *with values in* G; *under this pairing*

$$\mathcal{Q}^q(X, G) = \text{Annih } \mathcal{H}_q(X, I).$$

LEMMA 39.2. *Let* G *be discrete and* $\hat{G} = \text{Char } G$. *The Kronecker index establishes a dual pairing of* $\mathcal{H}^q(X, \hat{G})$ *and* $\mathcal{H}_q(X, G)$ *with values in the group* P *of reals* mod 1; *under this pairing*

$$\mathcal{H}_q(X, G) \cong \text{Char } \mathcal{H}^q(X, \hat{G})$$

$$\mathcal{P}_q(X, G) = \text{Annih } \mathcal{Q}^q(X, \hat{G}).$$

Both lemmas have been established for each of the complexes N_α. The passage to the limit is possible in view of formula (37.1.)

In $\mathcal{H}_q(X, G)$ we also consider the subgroup $\mathcal{T}_q(X, G)$ of all elements of finite

[28] For more detail see Lefschetz [7]. Although the definition of the homology and cohomology groups given here is valid for any space X, it is well known that its interest is restricted to compact spaces only. This is due to the fact that only in compact spaces is the family of finite open coverings cofinal with the family of all open coverings.

order. Since each approximating group $\mathcal{H}_q(N_\alpha, G)$ has a finite set of generators, one can show, by arguments resembling those of §24, that

$$\mathcal{T}_q(X, G) = \underrightarrow{\mathrm{Lim}}\ \mathcal{T}_q(N_\alpha, G).$$

40. Formulas for a general space

Using the formulas for complexes and applying a straightforward passage to the limit we obtain here some relations for $\mathcal{H}^q(X, G)$ and $\mathcal{H}_q(X, G)$ in terms of the groups $\mathcal{H}_q(X, I)$ and $\mathcal{T}_{q+1}(X, I)$. The results are not as complete as in the case of a complex.

THEOREM 40.1. *For a space X and a generalized topological coefficient group G the subgroup $\mathcal{Q}^q$ of the Čech homology group is expressible, in terms of a co-torsion group, as*

$$\mathcal{Q}^q(X, G) \cong \mathrm{Ext}^* \{G, \mathcal{T}_{q+1}(X, I)\}, \tag{40.1}$$

while the corresponding factor group $\mathcal{H}^q(X, G)/\mathcal{Q}^q(X, G)$ is isomorphic to a subgroup of Hom $\{\mathcal{H}_q(X, I), G\}$.

If G/mG is compact and topological for $m = 2, 3, \cdots$ then

$$\mathcal{H}^q(X, G)/\mathcal{Q}^q(X, G) \cong \mathrm{Hom}\ \{\mathcal{H}_q(X, I), G\}. \tag{40.2}$$

PROOF. For each nerve N_α we have (Theorem 32.1)

$$Q^q(N_\alpha, G) \cong \mathrm{Ext}\ \{G, \mathcal{T}_{q+1}(N_\alpha, I)\}$$

The groups on either side form inverse systems and it follows from Theorem 38.1 and Lemma 20.2 that the limits of these systems are isomorphic,

$$\mathcal{Q}^q(X, G) \cong \underleftarrow{\mathrm{Lim}}\ \mathrm{Ext}\ \{G, \mathcal{T}_{q+1}(N_\alpha, I)\}.$$

However since $\mathcal{T}_{q+1}(X, I) = \underrightarrow{\mathrm{Lim}}\ \mathcal{T}_{q+1}(N_\alpha, I)$ and the groups $\mathcal{T}_{q+1}(N_\alpha, I)$ are finite it follows from Theorem 24.2 that the limit on the right is Ext* $\{G, \mathcal{T}_{q+1}\}$. This proves formula (40.1).

From Theorem 32.1 we also have

$$H^q(N_\alpha, G)/Q^q(N_\alpha, G) \cong \mathrm{Hom}\ \{\mathcal{H}_q(N_\alpha, I), G\},$$

and again the limits of the two inverse systems are isomorphic in view of Theorem 38.1. Consequently from Theorem 21.1 we get

$$\underleftarrow{\mathrm{Lim}}\ [H^q(N_\alpha, G)/Q^q(N_\alpha, G)] \cong \mathrm{Hom}\ \{\mathcal{H}_q(X, I), G\}.$$

Now it follows from (20.1) (Chap. IV) that the group

$$\mathcal{H}^q(X, G)/\mathcal{Q}^q(X, C) = \underleftarrow{\mathrm{Lim}}\ H^q_\alpha/\underleftarrow{\mathrm{Lim}}\ Q^q_\alpha$$

is isomorphic with a subgroup of the group $\underleftarrow{\mathrm{Lim}}\ (H^q_\alpha/Q^q_\alpha)$. This proves the second assertion of the theorem. The subgroup will turn out to be the whole group whenever we are able to prove that $Q^q(N_\alpha, G)$ are compact topological groups.

Suppose now that G/mG is compact and topological for $m = 2, 3, \cdots$.

Given a cyclic group T of order $m \geqq 2$ we have Ext $\{G, T\} \cong G/mG$ (Corollary 11.2) and consequently Ext $\{G, T\}$ is compact and topological. It follows that Ext $\{G, T\}$ is compact and topological for every finite group T. In particular the groups

$$Q^q(N_\alpha, G) \cong \text{Ext } \{G, \mathcal{T}_{q+1}(N_\alpha, I)\}$$

are all compact and topological.

This completes the proof of the theorem. Notice that if G/mG is compact and topological for $m = 2, 3, \cdots$ then the group $\mathcal{Q}^q(X, G)$, as a limit of compact topological groups, is compact and topological.

If G is discrete and has no elements of finite order, or if $\mathcal{T}_{q+1}$ is countable, then by Theorem 24.4 and Corollary 24.1, the group Ext* in (40.1) may be replaced by Ext/Ext$_f$. In particular if $G = I$ then by Theorems 17.1 and 40.1,

$$\mathcal{Q}^q(X, I) \cong \text{Char } \mathcal{T}_{q+1}(X, I), \tag{40.3}$$

$$\mathcal{H}^q(X, I)/\mathcal{Q}^q(X, I) \cong \text{Hom } \{\mathcal{H}_q(X, I), I\}. \tag{40.4}$$

THEOREM 40.2. *The Čech homology group $\mathcal{H}^q(X, G)$ of a space X over a compact topological group G has a subgroup $\mathcal{Q}^q$, with factor group $\mathcal{H}^q/\mathcal{Q}^q$, both expressible in terms of integral cohomology groups of X as*

$$\mathcal{Q}^q(X, G) \cong \text{Char Hom } \{G, \mathcal{T}_{q+1}(X, I)\}, \tag{40.5}$$

$$\mathcal{H}^q(X, G)/\mathcal{Q}^q(X, G) \cong \text{Hom } \{\mathcal{H}_q(X, I), G\}. \tag{40.6}$$

PROOF. From Theorem 40.1 we have $\mathcal{Q}^q \cong \text{Ext}^* \{G, \mathcal{T}_{q+1}\}$. However since G is compact topological we have $\text{Ext}^* \{G, \mathcal{T}_{q+1}\} \cong \text{Ext } \{G, \mathcal{T}_{q+1}\}$ (Corollary 24.1) and $\text{Ext } \{G, \mathcal{T}_{q+1}\} \cong \text{Char Hom } \{G, \mathcal{T}_{q+1}\}$ (Theorem 15.1). This proves formula (40.5). We recall here that only continuous homomorphisms are considered. Formula (40.6) is a consequence of (40.2).

THEOREM 40.3. *The Čech cohomology groups $\mathcal{H}_q \supset \mathcal{P}_q$ of a space X over a discrete coefficient group G can be expressed, in part, in terms of the integral cohomology groups as*

$$\mathcal{P}_q(X, G) \cong G \circ \mathcal{H}_q(X, I), \tag{40.7}$$

$$\mathcal{H}_q(X, G)/\mathcal{P}_q(X, G) \cong \text{Hom } \{\text{Char } G, \mathcal{T}_{q+1}(X, I)\}. \tag{40.8}$$

PROOF. Let $\hat{G} = \text{Char } G$. Since $\mathcal{H}_q(X, G) \cong \text{Char } \mathcal{H}^q(X, \hat{G})$ and $\mathcal{P}_q =$ Annih $\mathcal{Q}^q(X, \hat{G})$ we have

$$\mathcal{P}_q(X, G) \cong \text{Char } [\mathcal{H}^q(X, \hat{G})/\mathcal{Q}^q(X, \hat{G})],$$

and using Theorems 40.2 and 18.1 we get

$$\mathcal{P}_q(X, G) \cong \text{Char Hom } \{\mathcal{H}_q(X, I), \text{Char } G\} \cong G \circ \mathcal{H}_q(X, I).$$

This formula could have been proved directly, passing to the limit with $\mathcal{P}_q(N_\alpha, G) \cong G \circ \mathcal{H}_q(N_\alpha, I)$. Since also $\mathcal{H}_q/\mathcal{P}_q \cong \text{Char } \mathcal{Q}^q(X, \text{Char } G)$, formula (40.8) is a consequence of Theorem 40.2.

The theorems and proofs carry over without change to the homology theory of X modulo a closed subset. Another generalization can be obtained by replacing the space X by a net of complexes, as defined by Lefschetz ([7] Ch. VI).

We are unable to answer the question whether $\mathfrak{Q}^q(X, G)$ and $\mathfrak{P}_q(X, G)$ are direct factors of $\mathcal{H}^q(X, G)$ and $\mathcal{H}_q(X, G)$. This is why we do not obtain expressions for $\mathcal{H}^q(X, G)$ and $\mathcal{H}_q(X, G)$ in terms of $\mathcal{H}_q(X, I)$ and $\mathcal{T}_{q+1}(X, I)$. The best we achieve in the case of a general space X is a description of the subgroups $\mathfrak{Q}^q$ and $\mathfrak{P}_q$ and of the corresponding factor groups, leaving the direct product proposition undecided.[29]

In the following sections of this chapter we shall discuss the case when X is a compact metric space, using the method of the fundamental complex. In this case we are able to obtain complete results, including the direct product decomposition.

41. The case $q = 0$

Before we proceed with the treatment of compact metric spaces we will discuss some details connected with the definition of the homology and cohomology groups for the dimension zero.

Let K be a finite simplicial complex. If we assume that there are no cells of dimension less than zero then every 0-chain will be a 0-cycle and the groups $H^0(K, G)$ and $\mathcal{H}_0(K, G)$ will be isomorphic to the product of G by itself n times, n being the number of components of K.

An alternate procedure is to consider K "augmented" by a single (-1)-cell σ^{-1} such that $[\sigma_i^0:\sigma^{-1}] = 1$ for all σ_i^0. In this case, given a 0-chain $c^0 = \sum g_i\sigma_i^0$, we have $\partial c^0 = (\sum g_i)\sigma^{-1}$ and consequently c^0 is a cycle if and only if $\sum g_i = 0$. The cohomology group gets affected also because the cocycle $\sum \sigma_i^0$ that was not a coboundary in the first approach is a coboundary in the augmented complex, since $\delta\sigma^{-1} = \sum \sigma_i^0$. It turns out that $H^0(K, G)$ and $\mathcal{H}_0(K, G)$ are isomorphic to the product of G by itself $n - 1$ times.

In defining the groups $\mathcal{H}^0(X, G)$ and $\mathcal{H}_0(X, G)$ for a space we again have two alternatives according as the nerves N_α are augmented or not.

Both the augmented and unaugmented complexes are abstract complexes in the sense of Ch. V and therefore all our previous results hold for either definition of $\mathcal{H}^0$ and $\mathcal{H}_0$. However in the discussion of compact metric spaces that follows there is an advantage in considering the nerves as augmented complexes, so as to have $\mathcal{H}^0(X, G) = \mathcal{H}_0(X, G) = 0$ if X is a connected space.

42. Fundamental complexes

Let X be a compact metric space. There is then a sequence U_n ($n = 0, 1, \cdots$) of finite open coverings of X such that U_n is a refinement of U_{n-1} and every finite

[29] Steenrod [9] §10 brings an argument, which if correct would settle the question positively. Unfortunately an error occurs on p. 681, line 5. The error was noticed by C. Chevalley, who has also constructed an example showing that the argument could not be corrected in the general case. If X is metric compact, Steenrod's argument can be corrected to give the desired direct product decomposition (see §44 below).

open covering of X has some U_n as a refinement. This last property asserts that in the directed family of all the finite open coverings of X the sequence $\{U_n\}$ constitutes a cofinal subfamily and therefore the Čech homology and cohomology group can be equivalently defined using only the sequence of coverings U_n. We shall assume that U_0 is a covering consisting of only one set, namely X itself, so that the nerve N_0 of U_0 is a vertex. For each n we select a projection $T_n : N_n \to N_{n-1}$ of the nerve of U_n into the nerve of U_{n-1}. The projections $N_n \to N_{n-k}$ we define by transitivity.

We now define the fundamental complex K of X as follows. The complexes N_n for $n = 0, 1, \cdots$ shall be disjoint subcomplexes of K. For each $n = 1, 2, \cdots$ and each simplex σ^q of N_n we introduce a new $(q + 1)$-cell $\mathfrak{D}\sigma^q$ whose boundary is $T_n\sigma^q - \sigma^q - \mathfrak{D}\partial\sigma^q$. This formula gives a recursive definition of the incidence numbers.

In order to give a more intuitive picture of K we may consider each of the nerves N_n as a geometric simplicial complex, the projection T_n can then be regarded as a continuous simplicial transformation; that is, as linear on every simplex σ^q of N_n, while $\mathfrak{D}\sigma^q$ can be visualized as a deformation prism consisting of intervals joining each point of σ^q with its image under T_n. With this interpretation K becomes a geometric complex and the cells $\mathfrak{D}\sigma^q$ can be subdivided so as to furnish a simplicial subdivision of K. It is clear from this picture that K can be contracted to a point, namely by moving every point up its projection lines towards the vertex N_0.

The complex K is countable and is locally finite; i.e., both closure and star finite. Viewing K as a closure finite complex, we can define finite cycles and infinite cocycles. However, since K is contractible all the homology group with finite cycles will vanish. Using the results of Ch. V we conclude that the cohomology groups with infinite cocycles also will vanish. Consequently, regarded as a closure finite complex, the structure of K is trivial. If we approach K as a star finite complex we obtain cohomology groups with finite cocycles and homology groups with infinite cycles. Regarded this way the complex K furnishes a true picture of the combinatorial structure of the space X.

43. Relations between a space and its fundamental complex

THEOREM 43.1. *The compact metric space X and its fundamental complex K are linked by isomorphisms*

$$\mathcal{H}^q(X, G) \cong H_w^{q+1}(K, G), \tag{43.1}$$

$$\mathcal{H}_q(X, G) \cong \mathcal{H}_{q+1}(K, G). \tag{43.2}$$

We shall restrict ourselves here to indicate the definitions of the isomorphisms without going into the complete proof, which involves lengthy but straightforward calculations.[30]

Let $\mathbf{z}^q$ be an element of $\mathcal{H}^q(X, G)$. Then $\mathbf{z}^q$ can be represented by a sequence

[30] This proof is closely related to one given by Steenrod; see [10], §4.

of cycles $z_n^q \in Z^q(N_n, G)$ such that $z_{n-1}^q - T_n z_n^q \in B^q(N_{n-1}, G)$. For each $n = 1, 2, \cdots$ select a chain c_{n-1}^{q+1} in N_{n-1} such that

$$\partial c_{n-1}^{q+1} = z_{n-1}^q - T_n z_n^q,$$

and consider the chain

$$z^{q+1} = \sum_{n=1}^{\infty} c_{n-1}^{q+1} + \sum_{n=1}^{\infty} \mathfrak{D} z_n^q.$$

We verify that

$$\begin{aligned} \partial z^{q+1} &= \sum_{n=1}^{\infty} (z_{n-1}^q - T_n z_n^q) + \sum_{n=1}^{\infty} (T_n z_n^q - z_n^q - \mathfrak{D}\partial z_n^q) \\ &= z_0^q - \sum_{n=1}^{\infty} \mathfrak{D}\partial z_n^q = 0, \end{aligned}$$

since $\partial z_n^q = 0$, while $z_0^q = 0$ for $q \geqq 0$, $z_0^0 = 0$ by §41. Consequently z^{q+1} is a cycle of K. If instead of $\{c_n^{q+1}\}$ we use a sequence $\{\bar{c}_n^{q+1}\}$ to define a cycle $\bar{z}^{q+1}$, then

$$z^{q+1} - \bar{z}^{q+1} = \sum_{n=1}^{\infty} (c_{n-1}^{q+1} - \bar{c}_{n-1}^{q+1})$$

Each term $c_{n-1}^{q+1} - \bar{c}_{n-1}^{q+1}$ is a finite cycle and therefore bounds in K, therefore $z^{q+1} - \bar{z}^{q+1}$ is a weakly bounding cycle and z^{q+1} determines uniquely an element $\mathbf{z}^{q+1} \in H_w^{q+1}(K, G)$. We define

$$\phi(\mathbf{z}^q) = \mathbf{z}^{q+1}.$$

Now let $\mathbf{w}^q \in \mathcal{H}_q(X, G)$. The element $\mathbf{w}^q$ can be represented for suitable n by a single cocycle $w^q \in \mathcal{Z}_q(N_n, G)$. We verify that $\mathfrak{D}w^q$ is then a $(q+1)$-cocycle of K. Using the formula

$$\delta w^q = \mathfrak{D} T_n^* w^q - \mathfrak{D} w^q \text{ in } K,$$

and the fact that $\mathfrak{D}$ and δ commute we show that $\mathfrak{D}w^q$ determines uniquely an element $\mathbf{w}^{q+1}$ of $\mathcal{H}_q(K, G)$. We define

$$\psi(\mathbf{w}^q) = \mathbf{w}^{q+1}.$$

We also notice that the pair of isomorphisms ϕ, ψ preserves the Kronecker index

$$\phi(\mathbf{z}^q) \cdot \psi(\mathbf{w}^q) = \mathbf{z}^q \cdot \mathbf{w}^q. \tag{43.3}$$

If X_0 is a closed subset of X then every covering U_n of X determines a covering of X_0 whose nerve L_n is a subcomplex of the nerve N_n of U_n. The subcomplex

$$L = \sum_{n=1}^{\infty} L_n + \sum_{n=1}^{\infty} \mathfrak{D} L_n$$

of K is then a fundamental complex of X_0. The isomorphisms (43.1) and (43.2) of Theorem 43.1 can be generalized as follows

$$\mathcal{H}^q(X \bmod X_0, G) \cong H_w^{q+1}(K \bmod L, G) \tag{43.1'}$$

$$\mathcal{H}_q(X - X_0, G) \cong \mathcal{H}_{q+1}(K - L, G). \tag{43.2'}$$

44. Formulas for a compact metric space

Using the fundamental complex and the results of Ch. V we shall now establish theorems for a compact metric space quite analogous to the ones proved for a complex in Ch. V.

THEOREM 44.1. *The Čech homology groups of a compact metric space X over a generalized topological coefficient group G can be expressed in terms of the integral cohomology groups $\mathcal{H}_q = \mathcal{H}_q(X, I)$, $\mathcal{T}_{q+1} = \mathcal{T}_{q+1}(X, I)$ as*

$$\mathcal{H}^q(X, G) \cong \mathrm{Hom}\,\{\mathcal{H}_q, G\} \times (\mathrm{Ext}\,\{G, \mathcal{T}_{q+1}\}/\mathrm{Ext}_f\,\{G, \mathcal{T}_{q+1}\}).$$

More precisely, in terms of the subhomology group $\mathcal{Q}^q$ of (39.3) *we have*

$$\mathcal{Q}^q(X, G) \text{ is a direct factor of } \mathcal{H}^q(X, G), \tag{44.1}$$

$$\mathcal{Q}^q(X, G) \cong \mathrm{Ext}\,\{G, \mathcal{T}_{q+1}\}/\mathrm{Ext}_f\,\{G, \mathcal{T}_{q+1}\}, \tag{44.2}$$

$$\mathcal{H}^q(X, G)/\mathcal{Q}^q(X, G) \cong \mathrm{Hom}\,\{\mathcal{H}_q, G\}. \tag{44.3}$$

To prove the theorem we use the fact that the Kronecker intersection is preserved under the pair of isomorphisms ϕ, ψ of the previous section. Consequently, since

$$\mathcal{Q}^q(X, G) = \mathrm{Annih}\ \mathcal{H}_q(X, I) \text{ in } \mathcal{H}^q(X, G),$$
$$Q_w^{q+1}(K, G) = \mathrm{Annih}\ \mathcal{H}_{q+1}(K, I) \text{ in } H_w^{q+1}(K, G),$$

we have

$$\phi[\mathcal{Q}^q(X, G)] = Q_w^{q+1}(K, G),$$

and the theorem becomes a consequence of Theorems 43.1 and 32.2.

THEOREM 44.2. *The Čech cohomology groups of a compact metric space X with coefficients in a discrete group G can be expressed in terms of the integral cohomology groups $\mathcal{H}_q = \mathcal{H}_q(X, I)$, $\mathcal{T}_{q+1} = \mathcal{T}_{q+1}(X, I)$ as*

$$\mathcal{H}_q(X, G) \cong (G \circ \mathcal{H}_q) \times \mathrm{Hom}\,\{\mathrm{Char}\, G, \mathcal{T}_{q+1}\}.$$

More precisely, in terms of the subgroup $\mathcal{P}_q$ of (39.4), *we have*

$$\mathcal{P}_q(X, G) \text{ is a direct factor of } \mathcal{H}_q(X, G), \tag{44.4}$$

$$\mathcal{P}_q(X, G) \cong G \circ \mathcal{H}_q, \tag{44.5}$$

$$\mathcal{H}_q(X, G)/\mathcal{P}_q(X, G) \cong \mathrm{Hom}\,\{\mathrm{Char}\, G, \mathcal{T}_{q+1}\}. \tag{44.6}$$

To prove the theorem we notice that

$$\mathcal{P}_q(X, G) = \text{Annih } \mathcal{Q}^q(X, \text{Char } G) \text{ in } \mathcal{H}_q(X, G),$$
$$\mathcal{P}_{q+1}(K, G) = \text{Annih } Q_w^{q+1}(K, \text{Char } G) \text{ in } \mathcal{H}_{q+1}(K, G),$$

and therefore

$$\psi[\mathcal{P}_q(X, G)] = \mathcal{P}_{q+1}(K, G)$$

and the theorem becomes a consequence of Theorems 43.1 and 33.1.

All these results remain valid for the homologies of X modulo a closed subset.

We now proceed to compare the results obtained here for the metric compact case with the results of §40 concerning general spaces.

Statements (44.1) and (44.4) contain a positive solution for the direct product problem which is still unsolved for the general space. Formula (44.3) was proved in (40.2) for general spaces only under the additional condition that G/mG be compact and topological for $m = 2, 3, \cdots$. Formula (44.2) was proved for general spaces under the form

$$\mathcal{Q}^q(X, G) \cong \text{Ext}^* \{G, \mathcal{T}_{q+1}(X, I)\}$$

which is equivalent to (44.2) because

$$\text{Ext}^* \{G, T\} \cong \text{Ext } \{G, T\}/\text{Ext}_f \{G, T\}$$

for countable groups T with only elements of finite order (Theorem 24.4) and the group $\mathcal{T}_{q+1}(X, I) \cong \mathcal{T}_{q+2}(K, I)$ is countable for a compact metric X, since K is countable.

Formulas (44.5) and (44.6) coincide with the ones proved in Theorem 40.3 for a general space.

45. Regular cycles

Using the concept of a "regular cycle" Steenrod ([10]) has defined a new homology group $H^q(X, G)$ of "regular" cycles, for a compact metric space X. This group is useful especially in the case when X is a subset of the n-sphere S^n, because it provides information about the structure of the open set $S^n - X$.

Steenrod ([10], Theorem 7) has proved that if K denotes a fundamental complex of X then

$$H^q(X, G) \cong H^q(K, G). \tag{45.1}$$

From this, using Theorems 43.1 and 32.1 we derive the formula

$$H^q(X, G) \cong \text{Hom } \{\mathcal{H}_{q-1}(X, I), G\} \times \text{Ext } \{G, \mathcal{H}_q(X, I)\}, \tag{45.2}$$

for $q > 0$. This formula expresses $H^q(X, G)$ in terms of $\mathcal{H}_{q-1}(X, I)$ and $\mathcal{H}_q(X, I)$ and hence shows that, essentially, $H^q(X, G)$ is no *new* invariant.

Let us specialize formula (45.2), assuming that $q = 1$, and that X is connected. We have then $\mathcal{H}_0(X, I) = 0$ and therefore

$$H^1(X, G) \cong \text{Ext } \{G, \mathcal{H}_1(X, I)\}. \tag{45.3}$$

Let us further assume $G = I$ and that X is one of the solenoids Σ. Since Σ is a connected, compact abelian group we have $H^1(\Sigma, P) \cong \Sigma$ (Steenrod [9], Theorem 15) where P (Steenrod's $\mathfrak{X}$) is the group of reals mod 1. Further, since Char $I \cong P$ we have $H_1(\Sigma, I) \cong \text{Char } H^1(\Sigma, P) \cong \text{Char } \Sigma$. Hence finally

$$H^1(\Sigma, I) \cong \text{Ext } \{I, \text{ Char } \Sigma\}. \tag{45.4}$$

This group will be explicitly computed in Appendix B; it was the starting point of this investigation (see introduction).

Steenrod has defined a subgroup $\tilde{H}^q(X, G)$ of $H^q(X, G)$ by considering regular cycles that are sums of finite cycles. He has also proved that under the isomorphism (45.1) this group is mapped onto the subgroup $B^q_w(K, G)/B^q(K, G)$ of $H^q(K, G)$.

We shall now show that, for $q > 1$,

$$\tilde{H}^q(X, G) \cong \text{Ext}_f \{G, \mathcal{H}_q(X, I)\}. \tag{45.5}$$

$$H^q(X, G)/\tilde{H}^q(X, G) \cong \mathcal{H}^{q-1}(X, G). \tag{45.6}$$

In fact, from Theorems 31.3 and 43.1 we deduce that $B^q_w(K, G)/B^q(K, G) \cong \text{Ext}_f \{G, \mathcal{H}_{q+1}(K, I)\} \cong \text{Ext}_f \{G, \mathcal{H}_q(X, I)\}$. This proves (45.5). In order to prove (45.6) notice that $H^q(X, G)/\tilde{H}^q(X, G) \cong H^q(K, G)/[B^q_w(K, G)/B^q(K, G)] \cong H^q_w(K, G) \cong \mathcal{H}^{q-1}(X, G)$.

Formulas (45.5) and (45.6) provide a splitting of $H^q(X, G)$ different from the one used in (45.2). The isomorphism (45.6) was established by Steenrod [10], who has also shown that $\tilde{H}^q$ can be computed using G and $\mathcal{H}_q(X, I)$, without however getting the explicit formula (45.5).

From (45.5) we immediately deduce the theorem of Steenrod that $\tilde{H}^q(X, G) = 0$ and $H^q(X, G) \cong \mathcal{H}^{q-1}(X, G)$ whenever $\mathcal{H}_q(X, I)$ has a finite number of generators.

Appendix A. Coefficient Groups with Operators

In many topological investigations it is convenient to construct homology groups $H^q(K, G)$ in cases when G is not just a group, but a ring or even a field. More generally, G can be allowed to be a group with operators. We show here that our results extend unchanged to such cases, and in particular, that the resulting homology groups are still completely determined by the integral cohomology groups.

G is called a *group with operators* Ω if G is a generalized topological group, Ω a space, and if to each element $\omega \in \Omega$ and each $g \in G$ there is assigned an element $\omega g \in G$ (the result of operating on g with ω), in such wise that

(i) ωg is a continuous function of the pair (ω, g),

(ii) $\omega(g_1 + g_2) = \omega g_1 + \omega g_2 \qquad (g_1, g_2 \in G)$.

It then follows that each element ω determines a (continuous) homomorphism $g \to \omega g$ of G into G; however, distinct elements of Ω need not determine distinct

homomorphisms. The set Ω may have a discrete topology, or may even consist of just one operator ω.

If both G_1 and G_2 have operators Ω, a homomorphism (or isomorphism) ϕ of G_1 into G_2 is said to be *Ω-allowable* if $\phi[\omega g_1] = \omega[\phi g_1]$ for all $g_1 \in G_1$, $\omega \in \Omega$.

If G has operators Ω, a subgroup $S \subset G$ is said to be *allowable* if $\omega(S) \subset S$ for all $\omega \in \Omega$. The operators Ω may then be applied in natural fashion to the factor group G/S, by setting $\omega(g + S) = \omega g + S$. Then G/S is a group with operators Ω, and the natural homomorphism of G on G/S is allowable.[31]

If G is a group with operators Ω, the various groups introduced as functions of G in Chapters I–IV are also groups with operators. Specifically, let H be a discrete group, and for each $\theta \in \text{Hom}\,\{H, G\}$ define $\omega\theta$ as $[\omega\theta](h) = \omega[\theta(h)]$. Then $\omega\theta \in \text{Hom}\,\{H, G\}$, and

(A.1) $$\text{Hom}\,\{H, G\} \text{ has operators } \Omega.$$

Furthermore, if $H = F/R$, where F is free, the groups $\text{Hom}\,\{F \mid R, G\}$ and $\text{Hom}_f\,\{R, G; F\}$ are allowable subgroups of $\text{Hom}\,\{R, G\}$, so

(A.2) $$\text{Hom}\,\{R, G\}/\text{Hom}\,\{F \mid R, G\} \text{ has operators } \Omega.$$

Again, let f be a factor set of H in G, and define another factor set ωf by taking $[\omega f](h, k)$ as $\omega[f(h, k)]$. Then Ω becomes a space of operators for the group $\text{Fact}\,\{G, H\}$. Furthermore $\text{Trans}\,\{G, H\}$ is an allowable subgroup; therefore

(A.3) $$\text{Ext}\,\{G, H\} \text{ has operators } \Omega.$$

In similar fashion one concludes that $\text{Ext}_f\,\{G, H\}$ and Ext/Ext_f have operators Ω.

As another case, take $\phi \in \text{Hom}\,\{G, H\}$ and define a homomorphism $\omega\phi \in \text{Hom}\,\{G, H\}$ by setting $[\omega\phi](g) = \phi[\omega(g)]$ for each $g \in G$. If G is compact or discrete, one may show that $\omega\phi$ is a continuous function of ω and ϕ. In this case, and for any generalized topological group H,

(A.4) $$\text{Hom}\,\{G, H\} \text{ has operators } \Omega.$$

In particular, if G is discrete or compact,

(A.5) $$\text{Char}\, G \text{ has operators } \Omega.$$

Given these interpretations of all our basic groups as groups with operators, we next demonstrate that the various isomorphisms between these groups, as established in Chapters II–IV, are allowable. In particular, an inspection of the construction used to establish the fundamental Theorem 10.1 of Chapter II proves

(A.6) *The isomorphism*

$$\text{Ext}\,\{G, H\} \cong \text{Hom}\,\{R, G\}/\text{Hom}\,\{F \mid R, G\},$$

where $H = F/R$, F free, is allowable.

[31] Practically all the elementary formal facts about groups and homomorphisms apply to operator groups and allowable homomorphisms.

The same conclusion holds for the other isomorphisms stated in that theorem. Also, the isomorphism Ext $\{G, H\} \cong$ Char Hom $\{G, H\}$ established in Theorem 15.1 for compact topological G and discrete H is allowable. The proof of this fact depends essentially on showing that the "trace" used in that theorem has the commutation property,

$$t(\omega\theta, \phi) = t(\theta, \omega\phi), \text{ for any } \theta \in \text{Hom } \{R, G\}, \text{ and } \phi \in \text{Hom } \{G, H\}.$$

The allowability of the other isomorphisms in Chapters II–IV is similarly established. The proofs are closely analogous to the "naturality" proofs of §12, except that here the operators apply to G, while in §12 the operator T applied to H.

Now turn to the homology groups. Let c^q be a chain in the star finite complex K, with coefficients chosen in the group G with operators Ω. For each $\omega \in \Omega$, define

$$\omega(c^q) = \omega(\sum_i g_i\sigma_i^q) = \sum_i (\omega g_i)\sigma_i^q;$$

since the result is a chain, and since the requisite continuity holds, the group $C^q(K, G)$ of q-chains has operators Ω. Moreover, $\omega\partial = \partial\omega$, so that both $Z^q(K, G)$ and $B^q(K, G)$ are allowable subgroups of C^q. Therefore

(A.7) $$H^q(K, G) \text{ has operators } \Omega.$$

The essential tool in establishing the isomorphisms of Chap. V is the Kronecker index $c^q \cdot d^q$ for $d^q \in \mathcal{C}_q(K, I)$, $c^q \in C^q(K, G)$. We verify at once that

(A.8) $$\omega(c^q \cdot d^q) = (\omega c^q) \cdot d^q \qquad (\text{all } \omega \in \Omega).$$

Since the subgroup A^q of Z^q was defined as a certain annihilator under this Kronecker index (see (29.9)), it follows at once that A^q is an allowable subgroup of Z^q. Furthermore, the proof that A^q is a direct factor of C^q depended on a decomposition of $\mathcal{C}_q$ as a direct product $\mathcal{C}_q = \mathcal{Z}_q \times \mathcal{D}_q$, for a suitably chosen group $\mathcal{D}_q$. In the notation of Lemma 16.2, we then had, by means of the Kronecker index (see the proof of Theorem 30.3)

$$C^q \cong \text{Hom } \{\mathcal{C}_q, G\} \cong \text{Hom } \{\mathcal{C}_q, G; \mathcal{D}_q, 0\} \times \text{Hom } \{\mathcal{C}_q, G; \mathcal{Z}_q, 0\}.$$

On the right both factors are allowable subgroups, and the isomorphism to the direct product is allowable;[32] furthermore, the second factor is the one which corresponds to the subgroup A^q of C^q. Therefore C^q has a representation of the form $C^q = A^q \times D^q$, where D^q is an allowable subgroup, complementary to A^q. A similar decomposition holds for Z^q and thus for its factor group $H^q = Z^q/B^q$. In terms of the homology subgroup $Q^q = A^q/B^q$ determined by A^q, this proves

(A.9) *The isomorphism* $H^q \cong (H^q/Q^q) \times Q^q$ *is allowable.*

The further analysis of these two factors, as carried out in Chapter V, all depended on the Kronecker index. In view of the property (A.8) of this index,

[32] If A and B are two groups with operators Ω the direct product $A \times B$ has operators Ω defined by $\omega(a, b) = (\omega(a), \omega(b))$ for $\omega \in \Omega$.

and the property (A.6) of the basic group-extension theorem, we have

(A.10) *The isomorphisms*

$$H^q(K, G)/Q^q(K, G) \cong \text{Hom } \{\mathcal{H}_q, G\},$$

$$Q^q(K, G) \cong \text{Ext } \{G, \mathcal{H}_{q+1}\}$$

are allowable, as is the isomorphism $H^q \cong$ Hom $\times$ Ext, *obtained by combining* (A.9) *and* (A.10).

Similar remarks apply to the representation of the "weak" homology group H^q_w (Theorem 32.2), which is a factor group of H^q by an allowable subgroup. The same holds for the topologized homology group H^q_t (i.e., the isomorphisms of Theorem 34.2 are allowable), for in any topological group G with operators Ω, the continuity of the operators insures that the subgroup $\bar{0} \subset G$ is allowable (recall that $H^q_t = H^q/\bar{0}$).

Turn next to the analysis of the cohomology groups. The groups $\mathcal{C}_q(K, G)$, with G discrete, again have operators in Ω, under the natural definition. As in the case of the homology groups, we have

(A.11) $\mathcal{H}_q(K, G)$ has operators Ω.

The representation of these groups depended on duality; i.e., on the Kronecker index $c^q \cdot d^q$, for $c^q \in Z^q(K, \text{Char } G)$, $d^q \in Z^q(K, G)$. Given the various definitions of the effect of an operator ω, one shows easily that

$$(\omega c^q) \cdot d^q = c^q \cdot (\omega d^q) \qquad (\text{all } \omega \in \Omega).$$

From this formula one may deduce that the well known isomorphism $\mathcal{H}_q(K, G) \cong \text{Char } H^q(K, \text{Char } G)$ is allowable. Thence it follows that the isomorphisms of Theorem 33.1 representing $\mathcal{H}_q$ are allowable.

These considerations yield the following

ADDENDUM TO THE UNIVERSAL COEFFICIENT THEOREM. *If K is any star finite complex, G a group with operators Ω, then the homology groups of K (and, if G is discrete, the finite cohomology groups) with coefficients in G all have operators Ω. All these groups with their operators are determined by the group G (with its operators) and the cohomology groups of the finite integral cocycles of K.*

A similar discussion applies to the results of Chap. VI.

In many important cases the operators form a ring (or even a field). Let us assume then that Ω is a generalized topological ring; that is, a ring which is a generalized topological group under addition and in which the multiplication is continuous. Then G is called an *Ω-modulus* if G is a generalized topological group with operators Ω (i.e., conditions (i) and (ii) above hold) such that

(iii) $$(\omega_1\omega_2)g = \omega_1(\omega_2 g) \qquad (\text{for } \omega_i \in \Omega, g \in G),$$

(iv) $$(\omega_1 + \omega_2)g = \omega_1 g + \omega_2 g \qquad (\text{for } \omega_i \in \Omega, g \in G).$$

In other words, addition and multiplication of operators are determined in the natural fashion from G.

If the standard coefficient group G is now assumed to be an Ω-modulus, simple

arguments will show that all the groups with operators Ω as described above are in fact Ω-moduli. Since the basic isomorphisms are still Ω-allowable, we conclude that the addendum to the universal coefficient group theorem still holds in these circumstances.

It is sometimes convenient to use a set Ω of operators in which only the addition or only the multiplication of operators is defined. More generally, we may consider a space Ω in which only certain sums $\omega_1 + \omega_2$ and products $\omega_1\omega_2$ are defined (and continuous); we then require that conditions (iii) and (iv) above hold only when the terms $\omega_1\omega_2$ or $\omega_1 + \omega_2$ are defined. The derived groups satisfy similar assumptions, and the universal coefficient theorem still holds.

If the coefficient group G is locally compact, one can always take the operators to form a ring, for any such group G has its endomorphism ring Ω_G as a natural ring of operators. Specifically, Ω_G is the additive group Hom $\{G, G\}$ of endomorphisms of G, with its usual topology (§3), and the multiplication $\omega_1\omega_2$ of two endomorphisms is defined by (iii) above. The requisite continuity properties of $\omega_1\omega_2$ and ωg are readily established, in virtue of the local compactness of G. Furthermore, if Ω is any other space of operators on G, each $\omega \in \Omega$ determines uniquely an endomorphism $\bar{\omega} \in \Omega_G$ with $\bar{\omega}g = \omega g$ for each g. The correspondence $\omega \rightarrow \bar{\omega}$ is a continuous mapping of Ω into Ω_G which preserves whatever sums and products may be present in Ω (assumed to satisfy (iii) and (iv)). Thus, any group, derived from G, which is an Ω_G-modulus is also a group with operators Ω, and any isomorphism between groups which is Ω_G-allowable is Ω-allowable. This indicates, that, for locally compact groups, one may restrict attention to operators of the ring Ω_G.

The most useful case is that in which the coefficient group is a field F, which is its own ring of operators. In this case all the homology groups, groups of homomorphisms, etc., become F-moduli; that is, vector spaces over F.

All these remarks suggest the following rather negative conclusion: *although in many applications it is convenient to consider a homology theory over coefficients which form more than merely a group, no new topological invariants can be so obtained.*

Appendix B. Solenoids

Here we compute the one-dimensional homology group $H^1(\Sigma, I)$ of regular cycles for the solenoid[33] Σ, or the isomorphic group Ext $\{I, \text{Char } \Sigma\}$ (see (45.4)).

A solenoid is uniquely determined by a Steinitz G-number; that is, by a formal (infinite) product $G = \prod p_i^{e_i}$ of distinct primes with exponents e_i which are non-negative integers or ∞. Any such number G can be represented (in many ways) as a formal product $G = a_1a_2 \cdots a_n \cdots$ of ordinary integers a_i; if G is not an ordinary integer, we can take each $a_i \geqq 2$. Given such a representation of G, take replicas P_n of the additive group P of real numbers modulo 1, and let ϕ_n be the homomorphism which wraps P_n a_{n-1} times around P_{n-1}. The

[33] Solenoids were studied by L. Vietoris, Math. Annalen 97 (1927), p. 459, and more in detail by D. van Dantzig, Fundam. Math. 15 (1930), pp. 102–135. See also L. Pontrjagin [8], p. 171.

P_n then form an inverse system of groups, relative to the homomorphisms $\psi_{n+m,n} = \phi_{n+1} \cdots \phi_{n+m}$, and the solenoid Σ_G is defined as the limit $\Sigma_G = \varprojlim P_n$. Therefore Char $\Sigma_G = \varinjlim$ Char P_n, where the groups form a direct system under the dual correspondences ϕ_n^*. Here Char P_n is an isomorphic replica I_n of the additive group of integers, and ϕ_n^* maps I_n into I_{n+1} by multiplying each $x \in I_n$ by a_n. Therefore Char $\Sigma_G = \varinjlim I_n$ is a subgroup N_G of the additive group of rational numbers, consisting of all rationals of the form a/d_n, with a an integer and $d_n = a_1 \cdots a_{n-1}$. Alternatively, N_G consists of all rationals r/s with s a "divisor" of G; hence N_G and Σ_G are uniquely determined by G, and are independent of the representation $G = a_1a_2 \cdots a_n \cdots$.

A Steinitz G-number which is not an ordinary integer also determines a certain topological ring. Set $G = a_1a_2 \cdots a_n \cdots$, $d_n = a_1 \cdots a_{n-1}$. In the ring I of integers, introduce as neighborhoods of zero the sets (d_n) of all multiples of d_n. Since the intersection of all these (d_n) is the zero element of I, these neighborhoods make I a topological ring. It can be embedded in a unique fashion in a minimal complete topological ring $I_G \supset I$, so that every element of I_G is a limit of a sequence of integers, under the given topology. This is one of the b_r-adic rings introduced by D. van Dantzig.[34] The additive group of I_G can be alternatively described as a limit of an inverse sequence; specifically, the factor group $I/(d_{n+1})$ has a natural homomorphism into $I/(d_n)$, and the limit group is $I_G \cong \varprojlim I/(d_n)$. In the special case when $G = p^\infty$ is an infinite power of a prime p, I_G is the ordinary ring of p-adic integers.

THEOREM. *If G is any Steinitz G-number which is not an ordinary integer, Σ_G the corresponding solenoid, and I_G the corresponding complete ring containing the ring I of integers, then*

$$\text{(B.1)} \qquad \operatorname{Ext}\{I, \operatorname{Char} \Sigma_G\} \cong I_G/I.$$

PROOF. As above, Char Σ_G is a group N_G of rationals, generated by the numbers $r_n = 1/d_n$ with relations $a_nr_{n+1} = r_n$. Therefore N_G can be represented as F/R, where F is a free group with generators $z_1, z_2, \cdots$, and R the subgroup with generators $y_n = a_nz_{n+1} - z_n$, $n = 1, 2, \cdots$. By the fundamental theorem on group extensions

$$\text{(B.2)} \qquad \operatorname{Ext}\{I, \operatorname{Char} \Sigma_G\} \cong \operatorname{Hom}\{R, I\}/\operatorname{Hom}\{F \mid R, I\}.$$

Let $\theta \in \operatorname{Hom}\{R, I\}$ and set

$$x(\theta) = \lim_{n\to\infty} [\theta y_1 + d_2\theta y_2 + \cdots + d_n\theta y_n].$$

Then $x(\theta)$ is a well-defined element of I_G, and $\theta \to x(\theta)$ is a homomorphic mapping of Hom $\{R, I\}$ into I_G and thus, derivatively, into I_G/I. We assert that the kernel of the latter mapping is Hom $\{F \mid R, I\}$.

Assume first that $\theta \in \operatorname{Hom}\{F \mid R, I\}$, and let θ^* be an extension of θ to F. Then

$$\theta(y_1 + d_2y_2 + \cdots + d_ny_n) = -\theta^*z_1 + d_{n+1}\theta^*z_{n+1},$$

[34] Math. Annalen 107 (1932), pp. 587–626; Compositio Math. 2 (1935), pp. 201–223.

so that the limit $x(\theta)$ is $-\theta^* z_1$, which is an integer in I. Conversely, suppose that $x(\theta) \in I$, and set $x(\theta) = -c_1$. We then have

$$\theta y_1 + d_2 \theta y_2 + \cdots + d_n \theta y_n \equiv -c_1 \pmod{d_{n+1}}.$$

By successive applications of this condition we find integers c_n with $\theta y_n = d_n c_{n+1} - c_n$. The homomorphism $\theta^* z_n = c_n$ then provides an extension of θ to F, so that $\theta \in \text{Hom}\,\{F \mid R, I\}$.

Every element in I_G is a limit of integers, hence has the form Lim $[b_1 + d_2 b_2 + \cdots + d_n b_n]$; therefore $\theta \to x(\theta)$ is a mapping onto I_G. We thus have

$$\text{(B.3)} \qquad \text{Hom}\,\{R, I\}/\text{Hom}\,\{F \mid R, I\} \cong I_G/I.$$

The correspondence is topological, as one may readily verify that both (generalized topological) groups carry the trivial topology in which the only open sets are zero and the whole space. Thus (B.2) and (B.3) prove the isomorphism (B.1).

By cardinal number considerations, one shows that the group I_G/I is uncountable, hence not void. The formula (B.1) gives at once all the special properties of the homology group of the solenoid, as found by Steenrod [10] in his partial determination of this group.

The University of Michigan
Harvard University

Bibliography

[1] Alexandroff, P. and Hopf, H. *Topologie*, vol. I, Berlin, 1935.
[2] Baer, R. *Erweiterungen von Gruppen und ihren Isomorphismen*, Math. Zeit. 38 (1934), pp. 375–416.
[3] Čech, E. *Les groupes de Betti d'un complexe infini*, Fund. Math. 25 (1935), pp. 33–44.
[4] Eilenberg, S. *Cohomology and continuous mappings*, Ann. of Math. 41 (1940), pp. 231–251.
[5] Eilenberg, S. and Mac Lane, S. *Infinite Cycles and Homologies*, Proc. Nat. Acad. Sci. U. S. A. 27 (1941), pp. 535–539.
[6] Hall, M. *Group Rings and Extensions, I*, Ann. of Math. 39 (1938), pp. 220–234.
[7] Lefschetz, S. *Algebraic Topology*, Amer. Math. Soc. Colloquium Series, vol. 27, New York, 1942.
[8] Pontrjagin, L. *Topological Groups*, Princeton, 1939.
[9] Steenrod, N. E. *Universal Homology Groups*, Amer. Journ. of Math. 58 (1936), pp. 661–701.
[10] ———. *Regular Cycles of Compact Metric Spaces*, Ann. of Math. 41 (1940), pp. 833–851.
[11] Turing, A. M. *The Extensions of a Group*, Compositio Math. 5 (1938), pp. 357–367.
[12] Weil, A. *L'Integration dans les groupes topologiques et ses applications*, Actualites Sci. et Ind. No. 869, Paris, 1940.
[13] Whitney, H. *Tensor Products of Abelian Groups*. Duke Math. Journ. 4 (1938), pp. 495–528.
[14] ———. *On matrices of integers and combinatorial topology*, Duke Math. Journ. 3 (1937), pp. 35–45.
[15] Zassenhaus, H. *Lehrbuch der Gruppentheorie*, Hamburg. Math. Einzelschriften, 21, Leipzig, 1937.

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 43, NO. 1

JANUARY, 1942

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is $5.00 a volume (four numbers); the price of a single copy is $1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

TABLE OF CONTENTS

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 43, NO. 2

APRIL, 1942

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is $5.00 a volume (four numbers); the price of a single copy is $1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

CONSERVATION OF SCHOLARLY JOURNALS

The American Library Association created this last year the Committee on Aid to Libraries in War Areas, headed by John R. Russell, the Librarian of the University of Rochester. The Committee is faced with numerous serious problems and hopes that American scholars and scientists will be of considerable aid in the solution of one of these problems.

One of the most difficult tasks in library reconstruction after the first World War was that of completing foreign institutional sets of American scholarly, scientific, and technical periodicals. The attempt to avoid a duplication of that situation is now the concern of the Committee.

Many sets of journals will be broken by the financial inability of the institutions to renew subscriptions. As far as possible they will be completed from a stock of periodicals being purchased by the Committee. Many more will have been broken through mail difficulties and loss of shipments, while still other sets will have disappeared in the destruction of libraries. The size of the eventual demand is impossible to estimate, but requests received by the Committee already give evidence that it will be enormous.

With an imminent paper shortage attempts are being made to collect old periodicals for pulp. Fearing this possible reduction in the already limited supply of scholarly and scientific journals, the Committee hopes to enlist the cooperation of subscribers to this journal in preventing the sacrifice of this type of material to the pulp demand. It is scarcely necessary to mention the appreciation of foreign institutions and scholars for this activity.

Questions concerning the project or concerning the value of particular periodicals to the project should be directed to Wayne M. Hartwell, Executive Assistant to the Committee on Aid to Libraries in War Areas, Rush Rhees Library, University of Rochester, Rochester, New York.

Copies of the following memoirs can be obtained by addressing

for the United States, Canada, and South America, The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

An introduction to the theory of elliptic functions. By GOSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By PAUL ALEXANDROFF. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By RICHARD COURANT. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. DOUGLAS. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By FRÉDÉRIC RIESZ. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. CHEVALLEY. 25 pages. Price 50 cents.

TABLE OF CONTENTS

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 43, NO. 3

JULY, 1942

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is \$5.00 a volume (four numbers); the price of a single copy is \$1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

Copies of the following memoirs can be obtained by addressing *for the United States, Canada, and South America,* The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

An introduction to the theory of elliptic functions. By GOSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By PAUL ALEXANDROFF. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By RICHARD COURANT. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. DOUGLAS. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By FRÉDÉRIC RIESZ. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. CHEVALLEY. 25 pages. Price 50 cents.

TABLE OF CONTENTS

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

SECOND SERIES, VOL. 43, NO. 4
OCTOBER, 1942

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is $5.00 a volume (four numbers); the price of a single copy is $1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

Copies of the following memoirs can be obtained by addressing

for the United States, Canada, and South America, The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

An introduction to the theory of elliptic functions. By Gosta Mittag-Leffler. Authorized translation by Einar Hille. 81 pages. Price 90 cents.

On various conceptions of correlation. By F. M. Weida. 37 pages. Price 75 cents.

Contact transformations. By L. P. Eisenhart. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. Rider. 52 pages. Price $1.00.

Tauberian theorems. By Norbert Wiener. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By Paul Alexandroff. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By Richard Courant. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. Douglas. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By Frédéric Riesz. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. Chevalley. 25 pages. Price 50 cents.

TABLE OF CONTENTS

ß